Quaestiones Infinitae

Agential Self-consciousness

beyond conscious agency

Zelfbewustzijn van Actoren

voorbij bewust actorschap

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Utrecht op gezag van de rector magnificus, prof.dr. G.J. van der Zwaan, ingevolge het besluit van het college voor promoties in het openbaar te verdedigen op vrijdag 25 januari 2013 des ochtends te 10.30 uur

door

Gerrit Harm Bos

geboren op datum 4 oktober 1983
te Kampen

Promotor: Prof. dr. M. Düwell

Co-Promotor: Dr. T. Müller

Contents

Acknowledgements

If one thing is certain, then it is that writing this dissertation was a thoroughly social affair. Of course, I wrote my dissertation and I did so in relative seclusion. Functional and necessary as that sort of seclusion was, however, it made sense only insofar as I was part of a group. I am a husband and father (to be), a son and brother, and a friend, acquaintance and neighbour. Or, in the case of my cats, I do not know what. This group life has been indispensable for this dissertation.

I would like to thank:

My supervisors Marcus Düwell, Thomas Müller and Micha Werner. Micha supervised me for three years, then he left for a one-year sabbatical. Micha basically let me do what I was doing and never hesitated to be realistic. He pointed out both improvements for and problems in what I was doing. That, I submit, is what I needed from a supervisor. Thomas, who supervised me during the final year, has been stimulating, critical and pragmatic. Partially as a result of this, I have been able to finish my project in time.

Marcus has and will continue to provide a background of trust in my philosophical capacities. He offered me the opportunity to design my own project and never showed any sign of regret, even though for three (long) years he said he failed to understand what my project was about. This blend of supervisors was good.

My colleagues here at the Philosophy Department: whether in Colloquia meetings of the Practical Philosophy group or during lunch, we have had good conversations and discussions. Thomas Fossen: we have a similar ambition in philosophy, which is to create new and interesting conceptions. Wang Xiaowai, for the dinners and conversations we had (and hopefully will continue to have). Annemarie Kalis and Frederike Kaldewaij, for the discussions we had in relation to our individual research interests.

The Volkswagen Stiftung, for including me as a member of their European Platform for Life Sciences, Social Sciences and Humanities. Within this framework I met colleagues and organized an interdisciplinary workshop in Berlin with them called 'Personal and Shared Intentions'.

Michael Quante, for the possibility of working for a while in Münster, and NWO for financing this. In Münster I was able to discuss some parts of my project with Michael Quante, and attend his course on Marx. David Schweikard, for his reading class in which we closely studied Hegel's writings on self-consciousness.

The 'intervisiegroep', and especially Maarten Prot. As PhD students at the same stage of our projects but with different scientific backgrounds and interests, we coached each other quite well.

My father and mother, for raising me with a need to assess daily matters as substantial issues, i.e., in terms of their fundamental aspects.

My sister-in-law Elisabeth van Doornum for helping me out with designing the cover.

And finally, my other half Hanna van Doornum for supporting me in many ways. I appreciate your natural habit of occasionally replacing my philosophical approach with a more appreciative one. Thank you for the painting on the cover, which I take to be symbolic in this respect.

0 – General Introduction

0.0 Introduction

When writing this introduction, I was drinking coffee in order to stay alert. In addition, I was *aware* that I was drinking coffee to stay alert. That I was aware of my action seems to be more than just coincidence. Apparently, my drinking coffee to stay alert *depended* on my awareness that my drinking coffee is a way to induce my alertness. This dependence relation, in turn, seems to *mark* my capacity to consciously act for a reason. Having this sort of explicit '... in order to ...'-concept of one's action, i.e., what I will call 'agential self-consciousness', is quite common among human agents.[1] It is equally

---

[1] This concept of agential self-consciousness stands in some contrast to concepts of 'agential self-knowledge' (Schmid 2011; Gertler 2011b, 47), 'agentive self-awareness' (Bayne and Pacherie 2007), 'sense of agency' (Bayne 2011), 'self-agency' (Metzinger 2011) and 'consciousness of agency' (Gallagher 2007). My view of agential self-consciousness denotes an agent's (i) *conceptual* awareness that one is acting for a reason, and allows us to ask (ii) whether such awareness depends on introspection, self-observation, self-evaluation or self-authorization at all.

normal to assume that our agential self-consciousness marks a capacity to consciously act for a reason.

The thesis that a capacity for agential self-consciousness *marks* a capacity to consciously act for a reason expresses an *ordinary* conviction which remains relatively unchallenged in *systematic philosophy*. The former – supposedly – marks the latter capacity in the sense that the former is generally conceptualized and explained in terms of the latter. I will call this thesis 'the ordinary thesis about agential self-consciousness'. This thesis goes back a long way. It inspired and continues to inspire many descriptive and normative, religious and secular stories about humanity. Therefore, it is a paradigmatic key to the way in which we rationalize the life we lead amongst other life forms and natural entities (cf. Cassirer and Lukay 2006 in particular part I).

In this book I will *develop* the thesis that agential self-consciousness is *independent* of a capacity to consciously act for a reason. I will not prove the thesis, but only demonstrate that it finds support in respectable philosophical accounts of action, reasons and self-consciousness.

On some superficial reading of it, this thesis which I intend to develop is widely accepted. We would agree, for example, that a human agent may drink coffee to stay alert out of habit, i.e., be largely unconscious of his action. In addition, we would agree that he may, retrospectively, conceive that he was drinking coffee to stay alert. As a result, we could agree that his '... in order to ...' conception of his drinking coffee as well as his drinking coffee are – somehow – independent of a capacity to consciously act for a reason. Consequently, it seems fair to conclude that (some form of) agential self-consciousness is independent of a capacity to consciously act for a reason. However, if pushed, we would, in the same breath, emphasize that even habitually drinking coffee to stay alert *depends* on a capacity to consciously act for a reason, if not on an exercise of it. That is, we would argue that an agent's '... in order to ...'-conception of his action depends on a conception that he is an agent who, while drinking coffee, could have but did not exercise his capacity to consciously act for a reason. Consequently, we would conclude that agential self-consciousness depends on a capacity to consciously act for a reason.

In this book I will develop a far more radical and quite *extraordinary* reading of the thesis that agential self-consciousness is independent of a capacity to consciously act

for a reason. I will call this thesis the 'extraordinary thesis about agential self-consciousness'. On this reading, the thesis states that an agent's '... in order to ...'-conception of his *past*, *present* and *future* action is independent of a capacity to consciously act for a reason. When I say 'independent', I mean that there is no need to ascribe a capacity to consciously act for a reason to an agent merely on the ground that he conceives that he was/is/will be acting for a reason.

The upshot of my argument will be a conceptual framework in which some important philosophical concepts are rearranged so that in terms of these concepts we can appreciate new philosophical questions as well as the urgency to address them. From my standpoint, the most important implication would be that this conceptual framework would allow us to take a step back from and investigate our presumed capacity to consciously act for a reason. It would challenge us to clarify the conditions under which agential self-consciousness would mark such a capacity, the extent to which *our* agential self-consciousness depends on such a capacity, and the moral significance of self-consciousness. Let me mention three concrete domains for which this thesis would pose a challenge. *Firstly*, it requires us to scrutinize our ordinary practices of punishment, praise and blame. We punish, for example, an agent only if, and to the extent that, at the time of action, he knew or could have known that what he was doing was against the law (Morse 2008, 7ff; Morse 2007). More generally, we presume that an agent who knows that his action has bad consequences would be more culpable than if he had not known this. For example, given that he knows that drinking coffee causes him to have sleepless nights, we blame an agent for drinking coffee instead of juice or tea. However, if agential self-consciousness is independent of a capacity to consciously act for a reason, then why make agential self-consciousness a ground for praise or blame? It is important to be able to take this question seriously, because blaming oneself or another agent is something we need to do on proper grounds. There would be something appalling in maintaining irrational forms of (self-)contempt (cf. Dillon 1997).

Secondly, and more generally, the extraordinary thesis about agential self-consciousness would require us to scrutinize the common assumption that the possibility of legitimate action evaluation depends on an agent's responsibility for that action. Such an account takes the legitimacy of action evaluation to depend on a (*meta-*)*physical* fact about that agent and his action, i.e., whether he *deserves* praise or blame for his action or the consequences thereof. Did *he* perform the action, or did it merely

happen to him? Would the action be bad if he performed it? This 'merit-based' approach to praise and blame stands in some contrast to the more 'pedagogical' one, which I will discuss below.[2] According to the merit-based account, the question of whether we are right to condemn someone for drinking coffee if this causes him to have sleepless nights would depend on the question of whether *he is the one who* is causing himself to have sleepless nights. The question of whether *he* is doing it depends on whether *he* chose to do it *or* whether *his* motives caused that action. In short, the question is for the extent to which as a subject he was involved or could have been involved in his action. Concepts of choice and motives depend strongly on concepts of practical reasoning, which in turn tend to assume that our awareness of motives and reasons correspond with a capacity to choose or determine our action. In light of the extraordinary thesis about agential self-consciousness, we are challenged to reassess the extent to which the merit-based approach is required or sufficient. Moreover, the extraordinary thesis about agential self-consciousness might turn out to be incompatible with traditional concepts of responsibility, and invite us to adopt a *pedagogical* approach to praise and blame. On the pedagogical approach, we might conceive action evaluation as a *practical measure* through which we educate, control and constrain human agents so as to facilitate a practical sphere in which harmonious or good life is possible.

Thirdly, the extraordinary thesis would require us to take a fresh look at conceptions of morality and at what ethics could be if agential self-consciousness is independent of a capacity to consciously act for a reason. A specific issue would be whether, and if so in which sense, ethical standards address us as conscious agents to normatively to abide by them. A related question is whether we are to conceptualize the idea that 'full-blown' human agents have a moral standing, and more particularly whether our standing is superior to that of rudimentary or disabled human or animal beings. We typically assume that human agents are, by virtue of their capacity to consciously act for a

---

[2] This terminology is inspired by Eshleman, who distinguishes a *merit-based* and a *consequentialist* account of moral responsibility (Eshleman 2009). In contrast to Eshleman, I would prefer to speak of a *merit-based* and a *pedagogical* (i.e., not consequentialist) account of *praise* and *blame* (i.e., not of responsibility). My reason for doing so is that, to my mind, moral responsibility is a term that has unduly strong connotations with the merit-based approach. Of course, the pedagogical approach needs to distinguish between actions which an agent performs and things that merely happen to him. However, there is no inherent need to cash out this distinction in terms of responsibility. We might as well make it an issue of whether or not the agent performing a particular action *depends on whether his doing something is caused by an agential mechanism which belongs to him naturally and is responsive to* praise or blame.

reason, subject to moral norms. Moreover, we assume that they have, as conscious rational agents, a moral status over and above that of embryos and animal beings. However, if agential self-consciousness is independent of a capacity to consciously act for a reason, then are we right in making such assumptions about morality and moral status? This challenge puts ethical issues in a new light: why rationalize our ethical practices in terms of our agential self-consciousness if agential self-consciousness in general is independent of a capacity to consciously act for a reason? By the same token, why discuss the moral status of animals, fetuses, very young children or severely mentally disabled or 'normal' human agents in terms of conscious capacities for rational action? (See Newson 2007; Engelhardt Jr 1988; Beauchamp 1999; Bortolotti 2007; for discussion see Gallagher 1996; Gruen 2010 in particular the first section; DeGrazia 1999.)

What reason do we have to hold that our agential self-consciousness marks a capacity to consciously act for a reason? This question runs the risk of being ridiculed for being either trivial or absurd. Is it not clear that we consciously act for a reason, when we know that we act for a reason? Consequently, it seems evident that our agential self-consciousness *marks* a capacity to consciously act for a reason. By the same token, it seems absurd to doubt that we have a capacity to consciously act for a reason. However, the question I raise *neither* denies that *our* agential self-consciousness marks a capacity to consciously act for a reason *nor* doubts that we are capable of consciously acting for a reason. Rather, it requires us to scrutinize the following question: does the fact that we have agential self-consciousness *depend on the fact* that we are capable of consciously acting for a reason?

As mentioned before, I will develop 'the extraordinary thesis about agential self-consciousness'. In our practical life, it might seem utterly inappropriate to endorse this extraordinary thesis. In our ordinary rationalizations of the practices in which we are involved, we typically stress that we, in contrast to other animal and natural entities, are capable of consciously acting for a reason. From this we infer that human beings owe respect to each other, but are – within some confines – allowed to instrumentalize the rest of nature, nature's animals or their habitats (i.e., as long as doing so does not entail disrespect for a present generation of human beings). We rely on this idea to spell out whether we owe something to animals, unborn and newborn babies, (very) young children, (severely) ill or disabled human beings, people who live in terrible circumstances (such as extreme poverty or extreme climates), future generations,

criminals and psychopaths. We rely on the idea that human agents have a capacity to consciously act for a reason when we rationalize our western liberal, political, legal and pedagogical institutions which aspire to improve, regulate and develop social practices in which human individuals are free to have and develop their own convictions and make their own choices. As a result, the extraordinary thesis about agential self-consciousness appears to be *either* hostile to these common practices *or* a mere philosophical curiosity which is of no practical consequence at all.

Contrary to appearances, however, the extraordinary thesis about agential self-consciousness is neither an integral rejection of these practices nor a mere philosophical curiosity. It is a thesis about agential self-consciousness, not about a capacity to consciously act for a reason. It is certainly a thesis which denies that agential self-consciousness *as such* marks a capacity to consciously act for a reason. It leaves open, however, whether or not *our* agential self-consciousness – due to some special design – marks a capacity to consciously act for a reason. Therefore it is not an integral rejection of our practices, nor of our rationalizations of these. The extraordinary thesis, if developed, would definitely have critical potential. It would imply that we need to explain whether, and if so why and to what extent, our agential self-consciousness marks a capacity to consciously act for a reason. By implication, it would require us to scrutinize our rationalizations of our practices to track the extent to which we are to rationalize them on the assumption that we are capable of consciously acting for a reason. Therefore, the extraordinary thesis about agential self-consciousness is significant and not a mere philosophical curiosity.

The remainder of this introductory chapter is in four parts. In the *first* section I will explain that our ordinary characterizations and explanations of (i) the actions of animals and children and (ii) – to some extent – the actions of those who seem to act against their own judgment invite us to develop the extraordinary thesis about agential self-consciousness. In the *second* section I will argue that, far from being a confusion arising from conflated ordinary talk about agency and consciousness, that invitation can be reconstructed in light of aspects of respectable philosophical traditions. In the *third* section I will give a more formal statement about my research question, hypothesis and methodology. In the *fourth* and final section, I will outline my argument in this book.

0.1 Non-paradigmatic Rational Action and Agential Self-consciousness.

In this section I will argue that, although extraordinary in some senses, the need to consider developing the thesis that agential self-consciousness is independent of a capacity to consciously act for a reason is already implicit in ordinary discourse. That is to say, our ordinary explanations of the actions of animals and very young children, if we take them at face value, would imply a view on rational agency which leans considerably in the direction of the extraordinary thesis. In addition, our common conviction that we sometimes act against our own judgment implies a conceptual problem which we could dissolve on the ground of the extraordinary thesis about agential self-consciousness.

0.1.1 Animal Agency and Agency of Young Children

As already mentioned, we ordinarily assume that a human being is capable of acting rationally *because he is capable of consciously acting for a reason*. Of course, we stress that a human agent quite often acts independently of explicit awareness that he is acting for a reason. We could agree that he may act out of reflex, habit, emotion, subconscious motives or subconscious convictions. For example, he may start crying if he hears a sad story or a particular piece of music, or when slicing onions (for discussion see Pollard 2005; Mele 2004; Zhu and Thagard 2002; Hursthouse 1991). Some of these actions may not be rational, but others may be. *However,* we do endorse the fact that our rational agency is distinctively marked by our capacity to consciously act for a reason. That is to say, despite the possibility that some of our actions are merely bodily or emotional reflexes regulated subconsciously or even independent of consciousness at all, we emphasize that distinctively human actions depend either *implicitly* or *explicitly* on a capacity to consciously act for a reason.

We also explain actions of animals and very young children in terms of reasons. We say, for example, that a cat is moving to its bowl in order to get some food *or* that the baby is reaching towards the ball to get hold of it. And these explanations seem to work perfectly well, even though we generally assume – in contrast to adult human agents – that animals and babies do not have a capacity to consciously act for a reason.

This raises the following question: does a reason-explanation of an action (implicitly) explain that action in terms of the assumption that the agent who performs it has a capacity to consciously act for a reason?

There are *two* ways to go about answering this question. We can submit that a rationalizing explanation of an action depends

(i)   on the thesis that rational agency depends on a capacity to consciously act for a reason; or

(ii)   on the thesis that rational agency is independent of a capacity to consciously act for reason.

The first answer should be avoided if we are to take at face value our rationalizing explanations about the actions of babies, toddlers and animals. For even if we agreed that babies, toddlers and animals have some form of consciousness or self-consciousness on which their agency depends, it would be too far-fetched to assume that they have a capacity to consciously act for a reason (for animals see Allen 2011; Andrews 2011; Bermúdez 2007; for children see Matthews 2010; Newen and Vogeley 2003; Rochat 2011). We could, of course, argue that our ordinary rationalizing explanations of the actions of animals and small children are confused *or* even disagree that we ordinarily give such explanations for the behaviours of these agents. That is, we could argue that animals and very young human beings do not act for reasons, and hence that we should not explain their behaviour in terms of reasons. In other words, we could emphasize that reason-explanations of such actions are false or – at most – true as metaphors. But this approach would dogmatically dismiss the *conceptual* point at stake. For it would assume that those who do give reason-explanations for behaviours of animals and young children cannot explain that rational agency is independent of a capacity to consciously act for a reason.

The conceptual point at stake is supported by the fact that we ordinarily rationalize the actions of agents who – presumably – lack a capacity to consciously act for a reason. The question is whether or not the fact that we give such explanations can be supported by an account of rational agency. Given that we deny that animals and small children have a capacity to consciously act for a reason, the question would be whether or not we could develop the thesis that rational agency is independent of a capacity to consciously act for a reason. As mentioned before, this conceptual question can only be

answered in two ways. *Either* we hold that rational agency depends on a capacity to consciously act for a reason *or* we hold that rational agency is independent of a capacity to consciously act for a reason. If we hold the first view, we must prove that we are *logically* bound to hold that rational agency depends on a capacity to consciously act for a reason (for an interesting example see Davidson 1982). Proof of this would, implicitly, yield proof that the extraordinary thesis about agential self-consciousness is false: for, given that agential self-consciousness is a feature of a rational agent, this proof would imply that agential self-consciousness depends on a capacity to consciously act for a reason.

Instead of arguing that rational agency depends on a capacity to consciously act for a reason, we could endorse the idea that some forms – but not the human form – of rational agency are independent of a capacity to consciously act for a reason. However, arguing in this way would ambiguously address the question of whether rational agency depends on a capacity to consciously act for a reason. To reduce this ambiguity, we could distinguish between a 'generous' and a 'strict' concept of rational agency. According to a *generous* concept, rational action is independent of a capacity to consciously act for a reason; according to the *strict* concept, it depends on a capacity to consciously act for a reason. Based on this distinction, we could then argue that the generous concept applies to animal agents, and to small children and human agents, but that the strict one applies to our rational agency and to our rational agency only (for various way to do this see Steward 2009; Bermúdez 2007; Velleman 1992). From this we could infer that our rational agency depends on a capacity to consciously act for a reason. However, no such thing follows about our rational agency. For if we allow for a generous concept of rational agency, we not only allow for the possibility that rational agency is independent of a capacity to consciously act for a reason, but we also implicitly require a justification of the strict concept of rational agency. That is to say, in light of the generous concept we should ask ourselves on which ground we assume that our rational agency depends on a capacity to consciously act for a reason. Can't we rationalize our actions purely in terms of the generous concept?

This latter question forces us to dig deeply for an. It requires us to spell out whether or not we can explain our rational agency and, in particular, our capacity for conceiving that we are acting for reasons independently of assuming a capacity to consciously act for a reason. Hence, it would require us to find out whether or not we can develop the

extraordinary thesis about agential self-consciousness, i.e., that agential self-consciousness is independent of a capacity to consciously act for a reason.

0.1.2 Extraordinary Forms of Agential Self-consciousness

Sometimes self-conscious agents do what they agree constitutes or causes something which is, on balance, bad (morally or non-morally), i.e., worse than what it would constitute or cause if they did something else. I will refer to this form of action as 'constituting or causing what one assents to be bad'. For example, an agent may drink coffee in amounts which he knows cause him to have horribly sleepless nights, despite being aware that he would not cause such bad symptoms if he was drinking tea. Such actions are familiar to us all. They are called, for example, sinful actions or, in a more secular mood, weak-willed or even irrational actions.

We are used to explaining an agent's rational action in terms of his (implicit) assent that it is good to perform that action. However, we also hold that an agent who is 'constituting or causing what he assents to be bad' does something while he (implicitly) assents that an alternative course of action would have better effects. We could conceive this latter kind of action as 'acting against one's judgment', i.e., as a failure of agency (Kalis 2011). The failure, supposedly, is that the agent *fails* to act on his implicit assent that a particular course of action is, on balance, good.[3] The idea that such failures are possible poses an explanatory challenge to our ordinary explanations of rational action: how do we account for the possibility of such a failure if we generally explain an agent's rational action in terms of his implicit assent that his action is, on balance, good?

To neutralize this explanatory challenge, we might assume a condition due to which an agent is not motivated to φ by his implicit assent that it is good to φ. For example, we might hold that he is motivated by a psychological or behavioural condition which he has never controlled (e.g., a dissociative disorder or ADHD) or by a condition which he did control but does not control anymore (e.g., some form of addiction) or by his own choices (e.g., starting to smoke or robbing a shop). The primary challenge is in explaining that, in the absence of conditions he never controlled, an agent can freely act

---

[3] Davidson (Davidson 1970) might reject this, arguing that the failure is merely that the agent assents on the wrong principle, i.e., it is not that he is acting against his own assent. Although the agent assents that in light of all available reasons φ-ing is better than ψ-ing, he nevertheless assents that ψ-ing is better than φ-ing. Based on this, one could even argue that doing ψ is not an instance of acting against one's own judgment. I will sidestep this issue in the present context, and focus on cases in which the agent seems to act against his own judgment.

against his own judgment. To explain this possibility, we may take either of the following two strategies. We may take an *internalist* approach and assume that the agent's capacity to make practical judgments consists of several departments, so that he can make practical judgments in one department which are overruled by judgments made in other departments. Alternatively, we may take an *externalist* approach and postulate a distinction between his capacity to infer which action is good and his capacity to motivate himself (see Stroud 2009).

Although it is surely interesting to frame 'constituting or causing what one assents to be bad' as 'acting against one's judgment', we must appreciate that doing so leads to complications which we could avoid if we adopted the extraordinary thesis about agential self-consciousness. We could elaborate on an apparent two-fold difference between 'constituting or causing what one assents to be bad' and 'acting against one's own judgment'. Firstly, the former assent is not obviously an assent that one's action is bad, whereas the latter obviously is. For example, there is a sense in which assenting that drinking coffee causes a condition that is worse than the condition that would be caused if one was drinking tea is independent of assenting that drinking coffee is worse than drinking tea. Secondly, the former assent is not obviously dependent on an agent's capacity to act on it, whereas the latter assent depends on such a capacity. For example, an agent may assent that drinking tea would result in a better condition than drinking coffee, irrespective of whether or not in so judging he can refrain from drinking coffee.

In light of this two-fold difference, let us assume that acting against one's judgment is a special, contingent form of assenting that one is constituting or causing something which is bad. Consequently, we could resist the idea that there is something inherently *irrational* or defective in 'constituting or causing what one assents to be bad'. We could, for example explain the possibility of 'constituting or causing what one assents to be bad' in terms of an agent's capacity to voice conceptions of the good quite independently of a capacity to act in accordance with these concepts, let alone doing so consciously.

If it were possible to argue that rational agency is independent of a capacity to consciously act for a reason, then we would need to be reluctant to argue that someone who is 'constituting or causing what one assents to be bad' is going against his own judgment. If, in addition, we could explain that an agent's capacity to tell whether his action has good or bad effects is independent of a capacity to consciously act for a

reason, then we should be even more reluctant to claim his agential self-consciousness marks a capacity to consciously act for a reason. For instead of acting against his own judgment, an apparently irrational agent may merely be doing x and at the same time voicing the negative evaluative statements which he has learned to make about acts like *x*. It may simply be a mistake to assume – in the absence of extreme conditions such as psychological disorders or addiction – that an agent's capacity to voice these evaluative statements depends on his capacity to bring about the good.

0.2 The Extraordinary Thesis in Philosophy

In light of the above considerations about rational agency and self-consciousness, one might expect philosophers to have an answer to the following question: does agential self-consciousness depend on a capacity to consciously act for a reason? In this section, I will clarify that this question is by and large absent in philosophy – at least when it comes to discussing agents who have agential self-consciousness. This question is absent because traditionally the phenomenon of agential self-consciousness has been explained on the assumption that rational agency depends on consciousness and that self-consciousness is a feature of the mind or its states. As a result, it has been the default position to assume that agential self-consciousness depends on a capacity to consciously act for a reason. I will argue that this default position could be implicitly challenged if we combined more recent models of intentional action, reasons and self-consciousness.

Traditionally, philosophers discussed rational agency as a feature of *human* agency, which presumably depends on consciousness. So we find, basically, three competing theses about human rational agency: a broadly Aristotelian one, a Humean one and a Kantian one. The first thesis is that a human being is capable of doing good because he has or is capable of having a concept of the good. Reason is the power that is used to infer what one should do and to align one's conduct in accordance with that. The second thesis is that a human agent has subjective motivational states and the faculty of reason to regulate these motivational forces so as to achieve the object which fulfils these motives. The third thesis is that a human agent has a self-concept due to which he is normatively required to act rationally, i.e., respecting his self-conscious agency as such.

These theses might be compatible to some extent. For example, philosophers endorsing the first or third will typically also endorse the second. However, philosophers who

embrace the second may resist the first and the third. In addition, there is some competition between the first and third.

It was at the time that Hume and Kant introduced their models of rational agency that philosophers became systematically interested in the nature of self-consciousness. Self-consciousness has, typically, been conceived as a reflexive feature of mental states or subjectivity.

In light of the above, we can explain why agential self-consciousness is usually discussed in terms of a capacity to consciously act for a reason. Such self-consciousness belongs to and is about rational agency which – presumably – depends on a capacity to consciously act for a reason, and it is a form of consciousness which is – presumably – only possible as a reflexive conscious capacity.

This broad and, admittedly, rough explanation of why a discussion of the extraordinary thesis about agential self-consciousness is as yet absent in contemporary philosophical debate should confer a sense of why we human beings have never really tried to challenge the ordinary thesis about agential self-consciousness.

There have been philosophers who challenged the modernist idea that the capacity to consciously act for a reason is to be valued, protected and developed over and above anything else. This idea has been criticized on at least three scores. It has been argued that it, and linguistic capacities more generally, alienates us from nature and our body in disturbing ways (Adorno and Horkheimer 1989). It has also been argued that this idea is merely something by which a weak human life form exerts control over healthy life forms (Nietzsche 1999). Still others have argued that this modern idea makes sense only against the background of its history and social conditions (Taylor 1992). Although significant, these philosophical schemas do not question whether our agential self-consciousness depends on a capacity to consciously act for a reason: they only point to the risks and limitations of taking a capacity to consciously act for a reason as the sole source of value above anything else.

More recently, findings and claims by behavioural scientists have motivated a discussion about the thesis of whether or not we are conscious agents or whether there is a 'self' which regulates action (Wegner 2003; Metzinger 2011; Sie and Wouters 2009). However, this discussion is not informed by a positive self-understanding which could replace the idea that we are 'selves with a free will'. At any rate, these accounts

are not inspired by, nor do they aspire to discuss the question of whether or not agential self-consciousness depends on a capacity to consciously act for a reason.

Therefore there might, from the standpoint of a contemporary philosopher, be no good reason for developing the extraordinary thesis about agential self-consciousness: everyone assumes that we are capable of consciously acting for a reason, and those who do not fail to support their view by an alternative account of self-consciousness.

Arguing in this way, no matter how intuitively appealing, would have to be philosophically unsound for *three* reasons. *Firstly*, it would amount to suppressing the legitimate conceptual worries raised in the previous section. *Secondly*, the question of whether we can develop the extraordinary thesis about agential self-consciousness is relevant in philosophy, for a sustained negative answer to it would count as support for a fundamental assumption in our philosophical self-understanding. A sustained positive answer to it would, in contrast, challenge us to explore whether, and if so why and to what extent, we can assume that our agential self-consciousness marks a capacity to consciously act for a reason. In light of this we could discuss whether, and if so to what extent, developing and protecting these conditions should be the highest priority in human practices. However, if we cannot give such an account, we might discuss what we would lose and what we would gain if we gave up the idea that we are capable of consciously acting for a reason. *Thirdly*, and this will be the topic of the subsequent two subsections, we can combine philosophical concepts of action, reasons and self-consciousness in ways which seem consistent with the extraordinary thesis about agential self-consciousness.

0.2.1 Intentional Action and Reasons for Action

Anscombe characterizes intentional actions as those which give application to a why-question for reasons (Anscombe 2000). For example, my drinking coffee is intentional *if* it gives application to the question 'why is he/are you drinking coffee?' This characterization of intentional action is not in terms of a capacity to consciously act for a reason.

Alongside this characterization of intentional action, we find philosophers arguing that reasons for action are *facts*, *states of affairs* or *events*, i.e., not mental entities such as beliefs, desires or concepts of the good or the just (see Alvarez 2010; Stout 1996; Bittner 2001). For example, the reason why I am drinking coffee is the event – i.e., not

my belief – that I am getting tired. Alternatively, it may be because of the fact that I should drink coffee to stay alert – i.e., not my desire to stay alert.

Once combined, Anscombe's account of intentional action and the latter type of account of reasons would entail, for example, that my drinking coffee is intentional *iff* I am drinking coffee because I am getting tired or because I must drink it to stay alert. This allows us to raise the question of whether 'rational action' (i.e., action performed for a reason) depends on consciousness at all, let alone on a capacity to consciously act for a reason. Consequently, it can be taken as an invitation to ask whether we are bound to account for rational agency in terms of a capacity to consciously act for a reason.

Of course, to answer this question we first need to clarify more exactly the nature of reasons and the motivational mechanisms which enable an agent to act for them. From this, we should then determine whether or not rational agency depends on a capacity to consciously act for a reason. Clarifying this will be the main task in chapters 2–4 of this dissertation.

0.2.2 Self-consciousness

Wittgenstein and Mead suggested that self-consciousness depends first and foremost on a social capacity for language (Mead and Morris 1967; for reconstruction of Wittgenstein's account of self-consciousness see Tugendhat 1979). Wittgenstein argued that self-consciousness is consciousness of a public entity, which (i.e., consciousness) depends on the capacity to refer to it in terms of a public language. This view of mind has inspired several philosophers to claim that self-consciousness cannot be reduced to features or faculties of private consciousness. From this it has been argued, for example, that self-consciousness is possible only as a result of public interactions among agents (Habermas 1988a; Apel 1975).

On my reading, Mead's and Wittgenstein's claim that self-consciousness depends on public features of consciousness and a capacity to refer to them and ourselves allows for another, more radical, interpretation. In terms of Mead's account of the significant symbol, we might as well consider the idea that these latter capacities enable specific forms of consciousness, but are and remain unconscious capacities. That is to say, we might consider explaining self-consciousness in terms of specific *pre*-conscious capacities for socially meaningful interaction.

Of course, this leaves us with several questions. What could these pre-conscious capacities for socially meaningful interaction be? Could such capacities constitute (agential) self-consciousness independently of a capacity to consciously act for a reason? I will assess these questions in chapters 5 and 6 of this dissertation.

0.2.3 Conclusion

For present purposes it suffices to make three considerations. Firstly, that there are philosophical frameworks which allow us to raise the question of whether rational agency depends on a capacity to consciously act for a reason. Secondly, that there are philosophical accounts of self-consciousness in terms of which we can ask whether self-consciousness depends on a capacity to consciously act for a reason. Thirdly, combining these philosophical accounts invites us to explore the possibility that agential self-consciousness is independent of a capacity to consciously act for a reason.

0.3 Research Question, Hypothesis and Method

The central research question of my book is: *can we develop the extraordinary thesis about agential self-consciousness?* If this thesis were developed, then we would be challenged to compare and assess the thesis in relation to assumptions which we ordinarily make about agential self-consciousness. I have no ambition to argue that this thesis holds true, only to prove that in light of respectable philosophical concepts there is conceptual space for the extraordinary thesis. My ambition in raising this question is quite modest and will only be understood by those who recognize that it is a question about conceptual assumptions concerning rational agency and self-consciousness and, in particular, about how far these assumptions feed into our understanding of our agential self-consciousness. More specifically, I will prove that the extraordinary thesis finds support in existing philosophical concepts of rational agency and self-consciousness. This I can prove, irrespective of whether or not these philosophical concepts are ultimately coherent or true. Hence, my thesis does not concern the ultimate coherence or truth of the extraordinary thesis. It is first and foremost my attempt to draw attention to certain extraordinary implications of existing concepts of rational agency and self-consciousness, i.e., implications which – if true – would challenge us to reconsider our ordinary views on agential self-consciousness.

My thesis is not an attempt to model (rational) agency and self-consciousness in general. I am content with, on the one hand, concepts of agency according to which agency is independent of rationality and, on the other hand, concepts of self-

consciousness and conscious agency according to which it is independent of agential self-consciousness or conscious rational agency. What is more, the extraordinary thesis does not challenge the possibility that some forms of rational agency depend on reflective capacities. Instead my thesis concerns the need to discuss a *theoretical* assumption which we make in understanding ourselves: that agential self-consciousness marks a capacity to consciously act for a reason. This need follows, I claim, from respectable models of rational agency and self-consciousness.

To put it simply: I am neither trying to disprove the ordinary thesis about agential self-consciousness nor to prove the extraordinary thesis about agential self-consciousness. What I do is draw attention to extraordinary – and as yet unnoticed – implications of certain ordinary models of rational agency and self-consciousness. To be successful, I must coherently integrate existing models of agency, reasons and self-consciousness so that they allow us to conceptualize the extraordinary thesis.

My hypothesis is that we can outline a philosophical account of rational agency and agential self-consciousness from which the extraordinary thesis about agential self-consciousness follows logically.

To prove my hypothesis, my method will be one of borrowing, revising and combining particular concepts and lines of argument from traditional and contemporary philosophical discussions about rational agency and self-consciousness. I will be quite selective, i.e., I will only take what is of use for me to develop the extraordinary thesis about agential self-consciousness.

This method may seem somewhat biased: it only focuses on accounts of rational agency and self-consciousness which would be useful to develop the extraordinary thesis about agential self-consciousness. But although it is focused, it is *not* biased. My aim is *merely* to develop (i.e., not to prove) the extraordinary thesis about agential self-consciousness. Therefore, I must merely prove that it can be developed, not that it holds water against competing accounts. Given that I neither endorse nor assess the extraordinary thesis, I am free to ignore claims and discussions about rational agency and self-consciousness, which start from concepts of agency and consciousness that are incompatible with the extraordinary thesis. I must ignore these claims and discussions if I am initially to track and then continue to keep track of the concepts and lines of

arguments which are relevant for answering the question of whether we can develop the extraordinary thesis about agential self-consciousness.

0.4 Overview

The remainder of this dissertation is in seven chapters. Chapter **one** will provide an initial characterization of what *mere* agential self-consciousness could be, independent of a capacity to consciously act for a reason. I will introduce this initial characterization by means of an imaginary agent who is drinking coffee in order to stay alert, while he *thinks that* and is *thinking as if* he might not have a capacity to act in light of this awareness. Then, *secondly*, I will argue that an agent who thinks like that suspends his judgment about the truth value of two opposing theses: the *special dependence thesis* and the *dual capacity thesis*. The special dependence thesis has it that an agent's capacity for agential self-consciousness depends – in one way or another – on a capacity to consciously act for a reason. The dual capacity thesis has it that the agent's capacity for agential self-consciousness is his capacity to make explicit the reason for which he is acting, which is independent of a capacity to consciously act for a reason. The special dependence thesis would be the sole ground on which the extraordinary thesis could be denied. I will argue that the dual capacity thesis would rationalize mere agential self-consciousness, and as such would give way to the extraordinary thesis. Finally, through a brief discussion of Velleman's account of the possibility of practical reason, I will characterize two conditions which must be fulfilled to develop the dual capacity thesis (Velleman 2000; Velleman 1989). *Firstly*, that a capacity to act for reasons is independent of a capacity to consciously act for reasons. The task of chapters 2–4 will be to warrant this first condition. *Secondly*, that a capacity to make explicit that one is acting for a reason is independent of a capacity to consciously act for a reason. The task of chapters 4–6 will be to warrant this second condition.

The purpose of chapter **two** is to clarify that we need a 'mind-independent' account of rational action, i.e., one according to which reasons for action are not mental entities and acting for them does not require consciousness of them. Despite some important limitations, Anscombe's analysis of 'intentional action' can be taken to suggests a *mind-independent* account of rational agency (Anscombe 2000). Anscombe seems to hold that action is intentional *iff* it gives application to a why-question for reasons that address the agent who performs the action. This leaves entirely open whether or not agents must have a mind to act rationally. What is more, she also suggests that animal actions

give application to a why-question for reasons. This leaves us with the following question: can we explain a capacity to act for a reason that is independent of mental capacities or at least independent of a capacity to consciously act for a reason?

Chapter **three** develops the thesis that reasons are facts about an action and that acting for these facts is independent of an agent's consciousness of them. Partially inspired by Taylor's account of behaviour explanation, I will develop the claim that a why-question for reasons for action is given application by action x, which happens because x is a factor for something - s (Taylor 1964; Stout 1996). To develop this claim, we can assume that a reason for action x is the fact that x is a factor for s. On this view of reasons, we could argue that a reason-explanation of action explains an action in terms of an agent's 'motivational responsiveness' to reasons. I will develop the idea that such motivational responsiveness is mind-independent. Furthermore, I will argue that this account of rational action is generous, but strict enough to distinguish rational action from mere reflexes or functional events.

The aim of chapter **four** is to clarify that acting for a reason is independent of a capacity to make practical judgments. I will develop the claim that reasons for action are independent of norms *and* that acting for them is independent of normative standards for action. In this chapter I will draw from Bittner's idea that reason-explanations are independent of norms and are objective conditions which explain an action as a response to these conditions (Bittner 2001).

In chapter **five** I will develop the thesis that making explicit that one is acting for a reason is itself independent of a capacity to consciously act for a reason. I will draw from and radicalize Wittgenstein's thoughts to develop the idea that language ranges over public entities only, hence that self-concepts depend on that self and its capacities being public. From Mead's account I will develop the thesis that a capacity to make explicit that one is acting for a reason is independent of a capacity to consciously act for a reason.

In chapter **six** I will develop a philosophical account of agential self-consciousness according to which agential self-consciousness is independent of a capacity to consciously act for a reason. I will first develop the thesis that agential self-consciousness is an agent's *immediate knowledge* that he is motivationally responsive to reasons. Against the background of Tugendhat's Wittgensteinian account of

epistemic self-consciousness, I will then introduce the idea that agential self-knowledge is the effect on an agent's consciousness of his exercising his capacity to indicate himself linguistically and assert that he is motivationally responsive to reasons (Tugendhat 1979).

In light of this explanation, I will conclude in chapter **seven** that we can develop the extraordinary thesis that agential self-consciousness is independent of a capacity to consciously act for a reason.

1 – Mere Agential Self-consciousness?

1.0 Introduction

We are used to conceiving of an agent who is *conscious that* he is acting for a reason as someone who is *consciously* acting for a reason. For example, we would claim that an agent who is conscious that he is drinking coffee to stay alert is consciously drinking coffee to stay alert. More explicitly, we would conceive an agent's consciousness that he is acting for a reason, as a form of consciousness by which he controls his action.

There are some reasons we could give for resisting our tendency to conceive of agential self-consciousness primarily in these terms. Attributing agential self-consciousness from a third-person standpoint amounts to ascribing to an agent a first-person consciousness of acting for a reason. Although such an agent's first-person attributions of agential self-consciousness would differ from and depend on different capacities than the third-person attributions of self-consciousness, I will develop the thesis that it is one thing to attribute agential self-consciousness to an agent, but quite another to attribute to him a capacity to consciously act for a reason. In other words, I will develop

the idea that 'mere agential self-consciousness' is possible. *Mere agential self-consciousness* would be agential self-consciousness independent of a capacity to consciously act for a reason.

My purpose in this chapter is limited: it should *portray* an agent who is *thinking that* and *as if* he might not have a capacity to consciously act for a reason. As an example, this should give us a sense of what mere agential self-consciousness could be. In this respect, it will form the background for the subsequent chapters, in which I will develop the idea of mere agential self-consciousness more fully.

This chapter is in four parts. In the *first* part I will position a two-level approach to agential self-consciousness in relation to different philosophical 'one-level' accounts of consciousness, agency and the relation between these two. In the *second* part I will portray an agent who is drinking coffee to stay alert. I will develop the thesis that *imagining* mere agential self-consciousness would amount to ascribing to him an *immediate* '... in order to ...'-conception that he is acting for a reason *independently* of thinking *that* and *as if* he is capable of acting on such a conception. In the *third* part, I will introduce 'the dual capacity thesis', which will be the key to an account of mere agential self-consciousness. The *dual capacity thesis* states that an agent's *capacity* to conceive that he is acting for a reason is independent of a capacity to consciously act for a reason. *Fourthly*, and finally, I will argue that Velleman's account of practical reason suggests a conceptual framework which – supposing that it can be revised in two ways – would allow us to develop the dual capacity thesis. I will characterize the conditions under which these revisions would be successful, and therewith set the task for exploring these revisions in the subsequent chapters.

1.1 Consciousness and Agency

How do consciousness and agency relate? Does agency depend on consciousness? Some argue that it does, but disagree as to what such consciousness would involve (Bayne Forthcoming; Gallagher and Zahavi 2010). Others argue that minimal action could be fully guided by unconscious motor processes (Proust 2003). Does consciousness depend on agency? Some argue that thinking, inferring and imagining anything whatsoever should count as acting. John Locke, for example, considered the following to be actions of the mind: creating ideas, comparing ideas and developing abstract ideas (cf. Uzgalis 2010 par 2.2). We find something similar in the Kantian idea that thinking depends on spontaneous acts of judgment (cf. McDowell 1994). Kant even argued that

reason is in itself practical, in the sense that it entails a law for action (Kleingeld 2010; Timmermann 2010; Willaschek 1991). Philosophers like Fichte and Hegel, for example, argued that self-consciousness depends on certain forms of conscious interaction with other conscious agents (cf. Neuhouser 1990; Brandom 2007). Phenomenologists like Heidegger favoured the idea that consciousness is essentially practical because it is basically our concern with our existence in this world (cf. Olafson 1975). Then there are behaviourists, who argue that a mental state, such as a belief, entails – or in a stronger version reduces to – a disposition to act in certain ways (Stout 2006; cf. Graham 2010). Finally, it has been argued that causal thinking depends on practical capacities because causes are 'handles or devices for manipulating effects' (Woodward 2008).

For the purposes of developing the extraordinary thesis, I am concerned with both questions: does consciousness depend on agency *and* does agency depend on consciousness? My main interest in these questions is specific in two ways. I will be asking whether an agent's *self*-consciousness, if he has it, depends on *conscious* agency. More specifically, *I will be asking whether an agent's consciousness that he is acting for a reason depends on his capacity to consciously act for a reason.* To answer the question, I have to explore both (i) whether the object of his consciousness, i.e., *his acting for a reason*, depends on a capacity to consciously act for a reason and (ii) whether an agent's *consciousness that he is acting for a reason* depends on this capacity.

It is a bit difficult to locate this question in contemporary philosophical discussion, especially because philosophers are largely concerned with agency insofar as it – supposedly – depends on consciousness and with an agent's self-consciousness insofar as it – supposedly – depends on conscious agency. In *philosophy of mind* we tend to assume that self-consciousness or self-knowledge, in general, is one's consciousness of one's own mental states or subjectivity (cf. Gertler 2011a). By implication, we tend to proceed as if an *agent's* self-consciousness is his consciousness of *conscious*, practical states such as desires, intentions, reasons *or* conscious agency (cf. Paul 2011). We explain an agent's consciousness that he is acting for a reason – if he has it – in terms of conscious determinants of action. For example, he knows what he intends and what he does intentionally, say, because he can introspect his mind (cf. Schwitzgebel 2010), because his intentions have a qualitative dimension (Shoemaker 1988) or because he is consciously authorizing his intention (Moran 2004a).

In *philosophy of action* we assume that an agent is conscious only in order to explain his action. This leaves us with the overall impression that an agent's consciousness of his action and motives, if he has it, is essentially practical. His consciousness takes the form of beliefs and desires, emotions, decisions, intentions, commitments, plans or rudimentary versions of these, or reflective practical capacities (Davidson 2006a; Davidson 2006b; Bratman 1987; Paul 2009; Hursthouse 1991; Anscombe 1956; Korsgaard 2009). In other words, in philosophy of action there is not really a concept of consciousness which would allow us to conceptualize an agent's consciousness that he is acting independently of a capacity to consciously act.

In *practical philosophy* we take an interest in an agent's consciousness of his action or his motives on the assumption that it enables him to reflect on his action and control these actions reflectively. Practical philosophers tend to agree that human individuals can reflect practically because they can assess whether performing a course of action would yield the object of desire. Some argue that human agents can and should act in light of other (higher-order) desires they have (Frankfurt 1971); their rational self-interest (Gauthier 1987); empathy (Hume 1978 book III part I; cf. Agosta 2011); values which are constitutive for their identity as agents, such as 'strong evaluations' (Taylor 1985, 16–42); rules which are constitutive for practices and institutions on which their agency depends (Rawls 1955; Schapiro 2001; M. Thompson 2008); or in terms of principles of rational action (Kant 1999; Kant 2003). As a result, practical philosophers, like philosophers of mind and action, do not provide for a concept of agency or consciousness which entails a concept of mere agential consciousness.

My attempt to develop the extraordinary thesis is compatible with concepts of self-consciousness and rational agency which are prevalent in philosophy of mind and action and in practical philosophy. I mention these concepts, however, because they are dominant backgrounds which might make it difficult for us to even *imagine* what agential self-consciousness could be independent of a capacity to consciously act for a reason.

It is hard to imagine this, because in these contexts we tend to discuss an agent's self-consciousness in terms of practical mental states, in terms of mental 'causes' of action or as the key constituent of a capacity for practical reason. It is of course viable to assume that we have practical mental states. It is equally viable to explain certain forms

of action in terms of mental 'causes'. And, of course, the connection between self-consciousness and practical reflection seems natural enough.

No matter how plausible these assumptions, they warrant no dogmatic assumption that an agent's consciousness that he is acting for a reason depends on a capacity for conscious control. We philosophers might at least entertain the possibility that conscious control depends on a form of agential self-consciousness which itself is independent of a capacity for conscious control. As a result, we must not dogmatically exclude the possibility that agential self-consciousness is, at heart, independent of a capacity to consciously act for a reason.

As far as I know, philosophers of mind, philosophers of action *and* practical philosophers have never addressed the possibility that an agent's consciousness that he is acting is independent of his capacity to act consciously. I suggest that we should make up for that. *Philosophers of mind* should address the possibility that an agent's consciousness that he is acting might – at heart – merely be his consciousness of himself and his states, which – although independent of observation – is independent of reflexive features of practical mental states. *Philosophers of action* should ask whether an agent's consciousness that he is acting could be independent of ascribing him a capacity to consciously act. *Practical philosophers* should discuss whether, and if so why and to what extent, our consciousness of our actions marks a capacity to reflectively control our actions.

In my book I seek to develop the thesis that *agential self-consciousness*, i.e., an agent's consciousness that he is acting for a reason, is independent of a capacity to consciously act for a reason. More specifically, I will develop the thesis that an agent's *conceptual consciousness that he performs a specific action for a specific reason* is independent of a capacity to consciously perform an action for a reason. To illustrate my research subject, I will introduce it in contrast to Korsgaard's claim that self-consciousness is the key to conscious control.

> I believe that human beings differ from the other animals in an important way. We are self-conscious in a particular way: we are conscious of the grounds on which we act, and therefore we are in control of them. (Korsgaard 2009, 19)

In this passage, Korsgaard clearly endorses that we control the grounds on which we act *by virtue of our consciousness of them*. However, it is not clear whether she endorses

the *conceptual thesis* that any possible agent who conceives the grounds on which he acts controls them consciously. Or does she merely endorse a *factual thesis* that *we* conceive the grounds on which we act in a way which gives us conscious control over them? In other words, does *any agent's* self-consciousness or only *human* self-consciousness depend on a capacity for conscious control?

Logically speaking, the conceptual thesis would imply the factual, but the factual might not imply the conceptual. The factual would not imply the conceptual, for example, if human agents distinctively had – independent of and in addition to their self-consciousness – a capacity which enabled them to control their actions on the ground of their self-consciousness. If this were so, then an agent's consciousness of the grounds on which he is acting would be necessary, but insufficient, for a capacity to consciously control them.

Korsgaard's explanation of her claim does not decide between these two interpretations:

> When you are aware that you are tempted, say, to do a certain action because you are experiencing a certain desire, you can step back from that connection and reflect on it. You can ask whether you should do that action because of that desire, or because of the features that make it desirable. And if you decide that you should not, then you can refrain. This means that although there is a sense in which what a non-human animal does is up to her, the sense in which what you do is up to you is deeper. (Korsgaard 2009, 19)

In this passage, Korsgaard implies that *we* can act on *conscious states*, i.e., that the *grounds* on which we act can be conscious states such as (our experiencing) a certain desire. In addition, she claims that our *awareness* of a ground gives us reflective control over this ground. Should this be so? Should grounds of action be experiences at all? Should the grounds on which *we* act be experiences? Should any agent's awareness that he is acting on a ground give him control over it? Should *our* awareness that we act on a ground give *us* control over it?

Korsgaard, in both passages which I quoted above, leaves open whether the grounds on which an agent acts should be conscious states. She merely implies that *we* can act on such states to illustrate how *our* awareness of the grounds on which we act gives us control over them. This raises *two* questions: (i) does she account for *our* control over

the grounds on which we act on the assumption that these grounds are conscious states? Or (ii) does she account for our control over them merely on the assumption that we are conscious of them – irrespective of whether or not these grounds are themselves conscious states?

Various philosophers have endorsed that reasons are external to the mind (cf. Stout 2005; Alvarez 2010; Bittner 2001). They would disagree with Korsgaard that a ground for action should be a conscious state, or even a mental one. It would in that sense be wrong to say, for example, 'I am drinking coffee on the ground of (my experiencing) a desire to stay alert.' It would be more appropriate to say 'I am drinking on the ground of the fact that drinking coffee is a way to satisfy a desire which I experience'. Thus conceived, a ground of action is a fact, not a desire. It is also not a belief that something is a fact. If externalism about reasons is right, then Korsgaard confuses, I think, something quite similar to what Alvarez calls 'explanatory' vs 'justificatory' reasons (Alvarez 2009). Alvarez would argue that although we may explain an agent's drinking coffee in terms of the *fact* that he experiences a desire to stay alert *or* the *fact* that he believes in drinking coffee as a means to stay alert, we can only justify his action in terms of *facts*. Facts that justify my action are, typically, not facts about my mental condition. For example, the ground on which I act is a fact that acting like this is a way to satisfy my desire. Hence, there is no need to assume that grounds of action are experiences at all, not even that they are mental states.

In light of this distinction between explanatory and justificatory reasons, Korsgaard should argue that our conscious control depends on our *being aware* that we act on grounds, rather than on these grounds being conscious states. Should our awareness of the grounds on which we act give us control over them? Korsgaard could argue in two ways that it should. She could argue that our awareness of them gives us reflective control over these grounds *or* over the practical effects that such grounds have. In light of externalism about reasons, i.e. the view that reasons are external to the mind and distinct from an agent's conscious states, the *first response* could amount to the following. By reflecting on the grounds for action, an agent *either* reinforces or diminishes these grounds *or* explores whether there are other ways to act for them. For example, by reflecting on the fact that drinking coffee is a way to satisfy his desire to stay alert, an agent may get to know other things about staying alert or drinking coffee which affect his desire to stay alert. These things may cancel/reinforce his desire to stay alert and, consequently, the fact that drinking coffee is a way to satisfy his desire.

Alternatively, he may find that he has no reason to stay alert, say because of other desires he has. For example, that it would be better for him to sleep for a while and improve his general level of alertness. Consequently, by reflecting on these facts he may affect his desire and hence the grounds on which he is acting.

The *second response* would entail that an agent's awareness of the grounds on which he acts depends on a reflective capacity that determines on which grounds for action an agent acts. This capacity has traditionally been identified as an agent's capacity to assess his inclinations practically in light of his further desires, goals or values, or in light of rationality or other norms. As mentioned above, it has been quite usual to hold that the grounds on which we act are in the mind or are represented in the mind: they are desires and beliefs, subjective principles of action, pro-attitudes, reasons or intentions. However, there has been, and continues to be, a minority of philosophers who resist the idea that reasons are in the mind (cf. E. Thompson 2007; Alvarez 2010)*.* In addition, there have been debates in philosophy of mind about whether or not mental states are conscious, subconscious or even unconscious (Ryle 2008; cf. Weintraub 1987; Searle 1991; Carruthers 2011). Hence, it is questionable whether or not motives for action depend on consciousness.

Once combined, these discussions in philosophy of action and philosophy of mind allow us to raise two questions. Should the grounds on which we act depend on mental states? Should the grounds on which we act depend on consciousness? These two questions give a new twist to the one which I previously addressed to Korsgaard: if the grounds on which we act might not depend on consciousness, then why hold that an agent's consciousness of them gives him control over them? Might an agent's consciousness of the grounds on which he is acting be independent of a capacity to consciously act on the grounds?

In a sense, this discussion of Korsgaard's claim about self-consciousness in agents has brought us to something which resembles the question of whether agential self-consciousness depends on a capacity to consciously act for a reason. The main difference – if there is any – would be that grounds for action and reasons for action are not the same. However, irrespective of whether or not there is such a difference between grounds and reasons, we can see how Korsgaard operates with the same concept of self-consciousness and action explanation which makes it so difficult to conceive of the extraordinary thesis about agential self-consciousness. Our discussion

of Korsgaard should make us aware that we cannot *just* assume that (rational) agency depends on consciousness or that self-consciousness enables conscious control.

Korsgaard does not address the question of whether any form of an agent's self-consciousness depends on a capacity for conscious control. As a result of that, we cannot decide whether or not her claim about self-control in terms of self-consciousness is a *conceptual* one. In response to her claim, I will treat it as a plausible thesis about our agential self-consciousness. In fact I think that, if true, this thesis would have important moral implications which are not easy to deny (cf. Beyleveld and Bos 2009; Beyleveld 1991). Nevertheless, I think that we should not assume that any form of agential self-consciousness marks a capacity for conscious control, and hence we must seek to assess the thesis that it does in terms of more general theories of agency, reasons and self-consciousness. To do so, we should, first of all, investigate whether our consciousness of the grounds on which we act *depends* on a capacity to control our acting on these grounds. That is, we need to explore the dependence relations between agential self-consciousness as such and a capacity to consciously act for a reason. The latter, obviously, depends on the former. But does the former depend on the latter? If not, does this mean that we should explain the latter as merely a contingent configuration of the former, i.e., in terms of specific psychological or social conditions which are inessential to agential self-consciousness? Does rational agency depend on a capacity to consciously act for a reason? And irrespective of whether or not it does, does the capacity for agential self-consciousness depend on a capacity to consciously act for a reason?

I have called agential self-consciousness which would be *independent* of a capacity to consciously act for a reason, 'mere agential self-consciousness'. In the remainder of this chapter I will illustrate what *mere agential self-consciousness* could be, if it existed.

1.2 '… in order to …'-Concepts and Agential Self-consciousness

Let us imagine an agent – other than ourselves – who is drinking coffee for a reason. We may ask: 'Why is he drinking coffee?', and we may answer, 'He is drinking coffee in order to stay alert.' Of course, our answer may be wrong, e.g., he may be drinking coffee for another reason or for no reason at all. However, it seems safe to assume that *if he is drinking coffee for a reason, then* a conception that he does so would be an '… in order to …'-conception that he is drinking coffee: that he is drinking coffee in order to stay alert *or* in order to satisfy his desire for it *or* in order to …, *etc.*

We tend to explain purposive actions in terms of intentional states of an agent, and we typically assume that these states are part and parcel of the agent's mind. This holds true not only for causalist approaches like Davidson's, but also for non-causalist ones like Anscombe's. There are philosophers who would apply '…in order to …'-concepts to the actions of entities which lack a mind. Dennett, for example, would argue that we are justified in applying such concepts to mindless non-agents, i.e., for purposes of explanation and prediction (cf. Dennett 1989). Frankfurt would argue that '… in order to …' concepts apply to behaviours of entities, as long as these behaviours are to be explained in terms of an *agent's guidance* of that behaviour. He stresses, however, that such guidance might, but need not, involve 'higher faculties' (Frankfurt 1978).

All these approaches, however, explain purposive action in terms of conscious agency. Frankfurt leaves us wondering what agential guidance might be independent of such higher faculties. Taking Dennett's intentional stance is a matter of treating an entity *as if it were an agent with beliefs and desires.* Davidson's beliefs and desires are states of consciousness, as are Anscombian intentions. Consequently, we lack a concept of rational action according to which '…in order to…'-conceptions of actions are independent of the attributions of conscious agency. Would such a concept be possible?

In the introduction to this dissertation I have argued that in ordinary life we do rationalize the actions of animals, i.e., of agents to whom we do not ascribe a capacity to consciously act for a reason. For example, we would allow one another to say 'the cat is walking to its bowl *in order to* get some food', even though we would emphasize that the cat's behaviour is independent of a capacity to consciously act for a reason. Of course, we might emphatically stress that our '… in order to …'-conceptions of this animal behaviour are merely metaphorical: the cat's behaviour is to be explained on a par with a thermostat's heating a room, i.e., it is independent of a capacity to act for a reason. In addition, we may stress that '… in order to …'-conceptions of drinking coffee are literal, and hence drinking coffee is to be explained in terms of a capacity to consciously act for a reason.

Although it makes sense to say that coffee-drinking – in contrast to walking to a bowl – is to be explained in terms of a capacity to consciously act for a reason, it would need to follow from models of rational action whether or not the cat's behaviour is not rational just *because it is independent of a capacity to consciously act for a reason*. No matter what traditional concepts of rational agency entail, it would surely be philosophical

dogmatism to reject ordinary reason- explanations of animal behaviour on the ground that we philosophers are used to explaining rational agency in terms of a capacity to consciously act for a reason. To reject ordinary explanations on the ground of these philosophical ones, we need proof that rational agency depends on a capacity to consciously act for a reason. In the absence of such proof, we can safely entertain the possibility of rational agency, independent of conscious, practical capacities.

To imagine *mere* agential self-consciousness, we must imagine a rational agent who,

  (i)   like a human agent, *has* an '…in-order-to…'-concept of his action; and
  (ii) like a cat, *lacks* a capacity to consciously act for a reason.

It is difficult to imagine what mere agential self-consciousness could be, because as a form of 'immediate' self-knowledge we typically explain it as a feature of practical mental states or conscious agency (cf. Tugendhat 1979; Anscombe 2000; Setiya 2010; Paul 2009; Korsgaard 2009). Of course, an agent may conceive himself observationally just as he conceives other agents, i.e., by virtue of the same observational or inferential capacities (cf. Carruthers 2011; Ryle 2008; Pears 1965). On the face of it, such *mediate* self-consciousness is relatively independent of conscious agency. Nevertheless, agential self-consciousness is typically quite immediate*.* To account for immediate self-consciousness, we explain it as a feature of conscious, practical states. Therefore, we tend to think that an agent can only have *immediate* '... in order to ...'-concepts of his action if he has a capacity to consciously act for a reason.

There are two basic obstacles to imagining mere agential self-consciousness. The first problem is understanding how an agent can *immediately* know that he performs actions independently of his performing them consciously. The second problem is understanding how his capacity to act for a reason is independent of a capacity to do so consciously. Therefore the question is: can we imagine an agent who *immediately* conceives that he is *unconsciously* acting for a reason?

This question has two components: a 'subjective' and a more 'objective' one. The *objective* component concerns the question of whether such a self-concept is ultimately coherent. The *subjective* component is whether an agent can conceive of his actions independently of thinking that or as if he can consciously act for reasons. In the remainder of this chapter I will elaborate on this subjective component. The other chapters of this book are devoted to elaborating the objective component.

31

In one sense, it is quite easy to imagine mere agential self-consciousness. For example, we can fairly straightforwardly imagine that our coffee-drinking agent thinks:

> *Initial report:*   I am drinking coffee to stay alert.'

If he thinks this, then he leaves open (or at least implicit) whether or not he thinks that or is thinking as if he is capable to consciously act for a reason. We would, of course, *ordinarily* assume that he implies the following report on his action.

> *Ordinary report:*   I assent that coffee-drinking is good, in light of *my awareness* that my drinking coffee makes me stay alert.

The key to imagining mere agential self-consciousness lies in appreciating that, no matter how intuitive, such ordinary reports are implicit in 'I am drinking coffee to stay alert' only if we assume that our coffee-drinking agent thinks *that* or *as if* he is capable of consciously acting for a reason. Consequently, to imagine what mere agential self-consciousness could be, we should ask whether the agent can think, 'I am drinking coffee to stay alert' independently of thinking that or as if he is capable of consciously acting for a reason. We would get the following picture.

> *Extraordinary report:*   I am drinking coffee while I am aware that drinking coffee makes me stay alert.

This extraordinary report is, prima facie, entailed by but does not entail the ordinary one. We have now drafted the subjective component of mere agential self-consciousness.

One might object that this concept of mere agential self-consciousness is much too simplistic, i.e., it does not describe the richness of agential self-consciousness. It is, of course, likely that our coffee-drinker's consciousness that he is drinking coffee to stay alert is surrounded by a whole lot of other things that he conceives. *Firstly*, the agent is *also* aware of alternative courses of action and what he would bring about if he performed one of these – for example, that he would become quite grumpy if he stopped drinking coffee *or* that he would relieve his sister if he were to phone her right now. *Secondly*, he is *also* aware that reasons for these alternative courses of action would be of a different kind. For example, drinking coffee to stay alert would amount to

acting for a prudential reason, whereas calling his sister to relieve her would count as acting for an altruistic reason. *Thirdly*, the agent also conceives whether, and if so to what extent, his present course of action is compatible with performing any of these alternative courses of actions. For example, he conceives whether or not he can drink coffee for altruistic reasons, say, not telephone but visit his sister to drink coffee with her.

Such aspects are generally, at least implicitly, part of *our* agential self-consciousness. I doubt, however, whether we should hold that agential self-consciousness *as such* depends on them. My concern is that doing so would unnecessarily complicate the task at hand: to portray what mere agential self-consciousness could be, if it existed. However, even if we accept that agential self-consciousness has these features, then this would still leave open whether or not extraordinary reports imply ordinary ones. That is to say, none of these three aspects depends on an agent's thinking that or as if he is acting in light of his awareness that he is acting for a reason.

There is a crucial psychological difference between the agent of an ordinary report and the agent of an extraordinary one. Let us briefly reflect on this difference before I spell out the basic thesis which we would need in order to explain the possibility of extraordinary reports. Ordinary reports reveal an agent who thinks that and as if his capacity to conceive courses of action in terms of reasons *is a source of reasons and a requirement to consistently act for them*. He thinks that and as if, by virtue of this capacity, he is responsible for finding out what he has most reason to do *in order* to do it. As a result, he will be satisfied with himself for doing what he thinks he has a compelling reason to do, or blame himself – or even fail to identify with or understand himself – if he does not do what he thinks he has compelling reason to do. By contrast, extraordinary reports reveal an agent who leaves open whether he is capable of doing *and* whether he is required to do what he thinks would be the most rational thing. He is able to be content with himself, even if he does not do the most rational thing.

1.3 Dual Capacity Thesis vs Special Dependence Thesis

Extraordinary reports depict mere agential self-consciousness. To conceptualize such reports we would need to endorse the first and resist the latter of the two following opposing theses:

Dual capacity thesis:	an agent's capacity to conceive his *present* course of action in terms of reasons is independent of a capacity to consciously act for a reason.
Special dependence thesis:	an agent's capacity to conceive his *present* course of action in terms of reasons depends on his capacity to consciously perform that action for a reason.

An extraordinary report would not imply an ordinary one *iff* we can coherently hold the dual capacity thesis. As a result, developing the concept of mere agential self-consciousness would amount to developing the *dual capacity thesis*.

The formulations of both theses can be revised so that they incorporate an agent's consciousness of his past and future actions as well.

Dual capacity thesis$_{rev}$:	an agent's capacity to conceive his past, present or future action in terms of reasons is independent of a capacity to consciously act for a reason.
Special dependence thesis$_{rev}$:	an agent's capacity to conceive his present/future course of action in terms of reasons depends on his capacity to consciously perform that present/future action for a reason. His capacity to conceive his past action in terms of reasons, insofar as he exercised that capacity prior to or at the time of his past action, depended on his capacity to consciously perform his past action for a reason.

The dual capacity thesis proposes a radical departure from the special dependence thesis. Nevertheless, as I will emphasize time and time again, those who endorse the dual capacity thesis may consistently hold that they have a capacity to consciously act for a reason. In other words, contrary to first appearances, we may consistently endorse the dual capacity thesis and endorse that *our* agential self-consciousness *marks* a capacity to consciously act for a reason. The dual capacity thesis merely denies that agential self-consciousness *depends* on a capacity to consciously act for a reason.

In the introduction I indicated my ambition to develop the extraordinary thesis about agential self-consciousness. We now have a concept of mere agential self-consciousness which gives some body to this extraordinary thesis. To develop the extraordinary thesis about agential self-consciousness, we need to develop the dual capacity thesis in the light of philosophical concepts of rational agency and self-consciousness. Let us now characterize the criteria for doing so successfully.

1.4 A Rudimentary Version of the Dual Capacity Thesis?

To isolate the sort of conceptual scheme in which we can develop the dual capacity thesis, let us briefly turn to Velleman's philosophy of practical reason (Velleman 2000; Velleman 1992; Velleman 1989). Velleman argues that practical reason depends for its possibility on, but does not reduce to, two capacities: a capacity to act and a capacity to explain one's actions. Practical reason emerges from these capacities, however, *iff* they are combined with an agent's disposition to act in ways which he can understand.

At first sight, Velleman seems to employ a conceptual scheme which is *friendly* to the dual capacity thesis. That is to say, we might take him to imply that a capacity to *act for a reason* and a capacity to *explain one's action in terms of reasons* are independent of a disposition to act in ways an agent can understand, i.e., independent of a capacity to consciously act for a reason. This reading of Velleman would, however, be faulty for at least two reasons. *Firstly*, Velleman would explain action in terms of mental states such as beliefs and desires which cause and justify action. He is basically concerned with explaining the nature of human *intentional* action, arguing that it is not merely caused by those beliefs and desires which justify action *but by the subject of these beliefs and desires.* To explain the subject's involvement in intentional action, he draws a distinction between *intentional action* and *mere activity*. He would argue that the former is regulated by practical reason, i.e., a capacity to consciously act for a reason. Now, although he argues that mere activity is not regulated by such a capacity, he would still assume that mere activity is regulated by an agent's desires and beliefs. As a result, it is hard to say whether Velleman would allow for a form of *rational action* which is independent of (some form of) a capacity to consciously act for a reason, even if it were independent of what he would call 'practical reason'. *Secondly*, Velleman seems to explain an agent's capacity to understand his action as the very same capacity by which he forms beliefs that – combined with desires – cause him to act. Hence it is hard to see, how such a capacity for self-understanding could be independent of a

capacity to consciously act for a reason. As a result of this, to develop the dual capacity thesis we would need an account of this capacity which secures that *this capacity is independent of a capacity to consciously act for a reason*. Therefore, Velleman's conceptual framework poses *two* obstacles to the dual capacity thesis. I will briefly explain these two obstacles below and then characterize the conditions under which we could circumvent them.

The *first* obstacle is that Velleman's account of practical reason is his attempt to distinguish intentional action from mere activity in terms of how they are subjectively regulated. He develops his account to argue that intentional action is not merely caused by beliefs and desires, but by the subject of these beliefs and desires. However, this way of drawing the distinction suggests that an agent, to act for a reason, must have practical reason. Although I think we should agree with Velleman that we cannot explain acting for a reason purely in terms of mental states that cause and justify an action, I think it is going too far to suggest that we need to explain the difference between mere activity and rational action on the assumption that the latter depends on subjective regulation. What is more, to develop the dual capacity thesis we are to prove that acting for a reason is independent of regulation by the sort of mental states of which agential self-consciousness would form a particular type.

The *second* obstacle is Velleman's account of an agent's capacity to *conceive* his action in terms of reasons. His idea that actions are caused by beliefs and desires makes it hard to conceive how an agent's capacity for agential self-consciousness – if he has it – is independent of a capacity to act on what he believes about himself and the action which he performs. But irrespective of whether or not rational action is regulated by such states, it would be an equally challenging obstacle to explain that an agent's capacity to understand his own actions is independent of a capacity to consciously act for a reason. For does exercising this latter capacity not depend on the form of self-consciousness which marks a capacity for conscious action?

Nothing in Velleman's conceptual framework forces us to ascribe to a rational agent with a capacity to act for a reason *and* a capacity to make explicit that he is acting for a reason, which is the sort of disposition that, according to Velleman, enables practical reason. To prove that an agent might not even have a capacity to consciously act for a reason, we would need to explain (i) that acting for a reason is independent of a capacity to consciously act for a reason *and* (ii) that an agent's capacity to make explicit

that he is acting for a reason is independent of a capacity to consciously act for a reason. Consequently, to develop the dual capacity thesis it would suffice if we developed the following final formulation of it:

Dual capacity thesis$_{fin}$: an agent's capacity to act for a reason and his capacity to make explicit that he is acting for a reason are independent of a capacity to consciously act for a reason.

The upshot of this discussion of Velleman's account of the possibility of practical reason is the idea that developing the extraordinary thesis amounts to developing *two* theses: that rational action is independent of a capacity to consciously act for a reason and, more generally, of mental regulation, and that self-consciousness is the product of exercising unconscious capacities.

1.5 Conclusion

In this chapter I have set out the conditions for an account of mere agential self-consciousness. What I set out is somewhat at odds with our usual accounts of rational agency and self-consciousness. Therefore we must explore whether we can develop it. I have identified the dual capacity thesis as the key to an account of mere agential self-consciousness. So to develop the extraordinary thesis that agential self-consciousness is independent of a capacity to consciously act for a reason, we should explore and map ways to develop the dual capacity thesis. More specifically, we should explore which concepts would allow us to argue that a capacity to act for reasons might be independent of a capacity to consciously act for reasons (chapters 2–4). In addition, we will need to explore which concepts would allow us to argue that an agent's capacity to make explicit that he is acting for a reason might be independent of a capacity to consciously act for a reason (chapters 5–6).

2 – Towards a Generous Account of Rational Action

2.0 Introduction

This chapter will set the stage for a 'generous account' of rational agency. A *generous account* would enable us to *conceive* and *explain* that acting for a reason is independent of a capacity to consciously act for a reason. In other words, it would explain that applying '... in order to ...'-concepts to some action *is* independent of attributing to an agent a capacity to consciously act for a reason. Such an account would allow us to develop the first half of the dual capacity thesis: that a capacity to act for a reason is independent of a capacity to consciously act for a reason. Given that the dual capacity thesis concerns a particular class of rational agency, i.e., those rational agents who have agential self-consciousness, this generous account should allow us to argue that rational agency is 'mind-independent'. It would be mind-independent if, according to this account, reasons are external to the mind and acting for reasons is independent of a mental representation of them. A generous account should meet this requirement to reduce the risk of adopting with it a vocabulary in which agential self-consciousness

would be the sort of mental entity which, on a mind-dependent account of rational action, ordinarily 'causes' and justifies action.

This chapter consists of *four* sections. In the *first* section I distinguish a mind-independent approach from a mind-dependent one in philosophy of action. The *mind-dependent* approach would allow us to argue that rational agency is independent of a capacity to consciously act for a reason *if* we can explain rational action in terms of pre-reflective consciousness. This approach, although it is somewhat generous, would offer dim prospects for the dual capacity thesis. That is to say, it depends on a philosophy of mind which makes it hard to argue that an agent *who has agential self-consciousness* has it independently of a capacity to consciously act for a reason. In the second section I will argue that Anscombe's mind-dependent account of *intentional action* gravitates around her mind-independent *definition* of rational action. In the *third* section I will argue that Anscombe's mind-dependent account of intentional action is potentially unwarranted in light of her mind-independent definition of intentional action, and is quite possibly at odds with her claim that non-linguistic animals may act intentionally. I will argue that Anscombe's account of intentional action poses *two* challenges to the dual capacity thesis. In the *fourth* section I will prepare the ground for addressing the first challenge in the third and fourth chapters of this dissertation.

2.1 Mind-dependent or Mind-independent?

We distinguish happenings that are mere events from ones that are actions (cf. Wilson and Shpall 2012; Casati and Varzi 2010). For example, a leaf blowing in the wind is merely an event, not an action. Raising a coffee cup is an action, not merely an event. In addition, we distinguish explanations of happenings in terms of non-conscious causes from those which are in terms of mental regulation. For example, the leaf is blowing in the wind because air is moving from a high- to a low-pressure area. The coffee cup is rising because an agent desires to drink coffee and believes in raising the cup as a means to drinking coffee. Mere events, we say, have unconscious causes. Actions are regulated by an agent, and most of the time by his mind – be it a rudimentary, pre-reflective mind or a fully reflective one. As a result, we are inclined to identify events in terms of non-conscious causation, on the one hand, and actions in terms of mental regulation, on the other.

In the previous chapter, we have briefly discussed Velleman's distinction between *action* and *mere activity*. Velleman clarifies the difference between the two with the

example of a president who is expected to open a meeting (Velleman 2000, 3–4). Unaware of what he is saying, the president says, 'I hereby declare the sitting closed.' According to Velleman, the president *unconsciously* performs this action. It is caused by his desire to close the meeting and by his belief that saying, 'I hereby declare the sitting closed' is a means to close the meeting, but not by *him*. Because it is caused in this way, his saying it differs from mere events such as a leaf blowing or a knee-jerk reflex. However, it differs from things he does in full awareness that he is doing them, e.g., his saying 'Excuse me, *I* said that unintentionally.' Velleman concludes that this requires

> [...] us to define a category of ungoverned activities, distinct from mere happenings, on the one hand, and from autonomous actions, on the other. This category contains the things that one does rather than merely undergoes, but that one somehow fails to regulate in the manner that separates human action from merely motivated activity. The philosophy of action must therefore account for three categories of phenomena: mere happenings, mere activities and actions. (Velleman 2000, 4)

Velleman characterizes mere activity *positively* as being something which, in contrast to mere happenings, an agent does. In addition, he characterizes mere activity *negatively* as being something that, in contrast to autonomous action, an agent 'somehow *fails* to regulate in the manner that separates human action from merely motivated activity'. Velleman continues to argue that

> [t]he boundaries separating these categories [i.e., mere happenings, mere activities and actions] mark increments in the subject's involvement as the cause of his own behavior. [...] Mere activity is therefore a partial and imperfect exercise of a subject's capacity to make things happen: in one sense, the subject makes the activity happen; in another, it is made to happen despite him, or at least without his concurrence. Full-blooded human action occurs only when the subject's capacity to make things happens is exercised to its fullest extent. To study the nature of activity and action is thus to study two degrees in the exercise of a single capacity. (Velleman 2000, 4)

Velleman explains the distinction between mere activity and action as an increment in a *subject's* involvement as the cause of his own behaviour. It is noteworthy that Velleman does say 'subject' and not 'agent'. On Velleman's view, the agent of mere activity and

intentional action is a *subject*. An agent is acting, in the full sense, *if* his action depends on his subjectively concurring to it.

This distinction between mere activity and intentional action has (and here we have an eye on our attempt to develop a generous account of acting for a reason) *two* relevant features. Firstly, it suggests that mere activity is *purposive* in some sense. Hence, there is a sense in which a president's saying, 'I hereby declare the meeting closed' is '... in order to ...', yet is not consciously performed. Of course, Velleman would imply that this form of acting for a reason falls short of being the form of acting for a reason which marks human intentional agency. But this is immaterial to my attempt to develop the dual capacity thesis. Secondly, however, Velleman suggests that an agent engaged in mere activity is still involved as a subject – be it to a lesser degree than if he were acting intentionally.

Velleman's approach to intentional action stands in contrast to what he calls 'the standard story'. He argues that philosophy of action should, but standardly fails to, explain the distinctive sense in which a human agent can be said to act intentionally. The standard story about intentional action states that an agent's '[...] desire for the end, and his belief in the action as a means, justify taking the action, and they jointly cause an intention to take it, which in turn causes the corresponding movements of the agent's body [...]' (Velleman 1992, 461; cf. Mele 2004). Velleman's objection to this story is that it fails as an explanation of human intentional action: it fails to explain the distinctive sense in which a human *subject* regulates his intentional action (Velleman 1992, 462, 465). That is to say, it makes the subject a more or less passive bystander in relation to his beliefs and desires that are causing him to act. To explain intentional action, Velleman argues, we must agree with the standard story that human intentional action is caused by beliefs and desires, but, vitally, *also* by the agent's *disposition to act* in ways that he can understand. If we agree with this, then a human agent, in contrast to an animal agent, is subjectively involved in regulating his activity. Firstly, he does what he does only if and because he (implicitly or explicitly) understands his doing it in terms of his desires and beliefs. Secondly, a human agent will continue/start a behaviour if and because he can *predict* it from his beliefs and desires.

Velleman's claim that the standard story fails to explain the sense in which human agents seem to be involved in regulating their action is compatible with, but immaterial to, an attempt to develop a generous account of rational action. However, several

aspects of Velleman's account might stand in the way of developing such a generous account. Firstly, Velleman's terminology suggests that something qualifies as an 'action' only if an agent is subjectively regulating his activity in light of what he can understand on the ground of his beliefs and desires. Although, of course, we can agree to use the terms in this way, this should not blind us to the possibility that doing something for a reason might not depend on subjective regulation at all (cf. Hurley 2003; Heyes and Dickinson 2007; Steward 2009; Boyle 2003; Bermúdez 2007). Frankfurt, for example, argued that a spider can be said to act, i.e., to *guide* its behaviour. Frankfurt's basic idea is that the spider, although it lacks higher cognitive capacities, seems to have a capacity to compensate for events which distort what it is doing. For example, if the spider walks in a straight line and I put my finger in front of it, it will be able to avoid my finger and then continue in a straight line (cf. Frankfurt 1978). Secondly, Velleman's account of acting for a reason relies heavily on the assumption that the difference between mere activity and human intentional action marks an increment of *subjective* regulation. As a result of this, Velleman misses out on a chance to clarify the difference in terms of another obvious difference between the two. In particular, he does not pay enough attention to the fact that mere activity is typically quite *accidental*, whereas action is much more *diligent*. Frankfurt's example of the spider illustrates a case of unreflective diligence. Therefore the difference between acting for a reason and mere activity might be put as a difference between accident and diligence, irrespective of whether or not the agent regulating it has a mind.

Of course, we can doubt that the unreflective spider acts for a reason. However, what about African vultures that – after years of self-education – are able to grab a bone which they cannot swallow whole, fly above a mountain and release the bone so that it hits the rocky mountain top? They do this until the bone is smashed into pieces and they can swallow it (Anon. 2008a). This sort of behaviour differs from Velleman's examples of mere activity, because it is quite *diligent*. It differs from the spider walking, because it appears to be structured in a much more complex way in that it is continued until, through it, a particular state is realized. It differs from human intentional action because, although both are performed diligently, the vulture's behaviour is – presumably – independent of a capacity to consciously act for a reason.

This would suggest that besides *mere happenings*, *mere activity* and *human intentional action* there might be another form of movement we can call 'rational action'. Or, how

are we to describe such diligent and seemingly purposive movement? As intentional action? As non-intentional mere activity? As mere happening?

Remember that I am primarily dealing with the question of whether or not acting for a reason depends on a capacity to consciously act for a reason. I am not interested in questions about what the nature of intentional action is or whether such diligent behaviour qualifies as intentional action. In short, it does not really matter in the present context whether Velleman's or the standard story gives the right account of intentional action. I mention them merely for the prospects they leave for a generous account of rational action. The standard story and Velleman's offer us two ways to develop a generous account of acting for a reason. We could, in line with the standard story, argue that to be capable of acting for a reason an entity must, if he lacks a capacity to consciously act for a reason, at least have mental states which justify and *persistently* regulate his activity. Alternatively, we could argue in a Vellemanian mood that to act for a reason an entity must, if he lacks a capacity to consciously act for a reason, have a pre-reflective capacity to agree to his activity in light of his pre-reflective mental states which cause and justify his action.

Both of these two options have a significant, if not fatal, drawback. They are '*mind-dependent'* accounts of agency, i.e., ones in which a capacity to act for a reason is a mental one. Reasons are, presumably, mental states which cause and justify an action. Presumably, such states can be pre-reflective, rudimentary versions of the full-blooded ones which – supposedly – mark human action. Hence, although such an account would allow us to argue that a capacity to act for a reason is independent of a capacity to consciously act for a reason, it would do so only *to the extent that the agent to which it belongs lacks conceptual self-consciousness*.

However, to develop the second part of the dual capacity thesis we would need to develop the idea of an unconscious capacity to act for a reason, on the assumption that the agent who has it has *agential self-consciousness*.

The problem with a 'mind-dependent', generous account of rational action is that it conceives and explains action in terms of mental causation or regulation. No matter how 'light' this is, i.e., rudimentary and pre-conceptual, it comes with an ontology of reasons or acting for reasons which makes it hard to explain that agential self-consciousness is independent of a capacity to consciously act for a reason. Therefore,

without rejecting a mind-dependent, generous account of rational action, we have to sidestep it and develop a mind-independent, generous account of rational action. If successful, this *mind-independent account* would allow us to claim that reasons need not be in the mind and that acting for reasons is independent of a mental representation of them.

Anscombe's account of intentional action offers an almost perfect starting point for a mind-independent, generous account of rational action. Let us therefore discuss it in some detail (Anscombe 1979; Anscombe 2000).

2.2 Anscombe: Mind-independent Rational Action?

Anscombe claims that an action is intentional *if* it gives application to a why-question for reasons (Anscombe 2000 par 5). She distinguishes a why-question for *reasons* from a why-question for *causes*. An action gives application to a why-question for causes if that action can be explained as the effect of some cause. An action gives application to a why-question for reasons if it can be explained in terms of the agent's intention with it. Consider, for example, an agent who is drinking coffee to stay alert. Why does he drink coffee? Because he *desires* to stay alert and *believes* in drinking coffee as a means to stay alert. This would be a causal explanation. However, there is a form of why-question which we can answer by citing the fact that the agent is trying to stay alert.

Anscombe's definition of intentional action (as action which gives application to a why-question for reasons) is at first sight *mind-independent*, i.e., independent of the mental features of an agent who does the action. That is to say, it does not assume that the why-question applies only to actions which are mentally regulated. It is easy to overlook the fact that this definition is mind-independent, for at least three reasons. *Firstly*, we tend to assume that reasons are mental entities or that acting for a reason depends on mental or cognitive capacities such a practical reason (cf. Millgram 2008). *Secondly*, there is Davidson's influential idea that reasons for action reduce to mental causes which justify the action (Davidson 2006a). *Finally*, Anscombe herself develops her definition of intentional action in terms of a mind-dependent account of intentional action.

Nevertheless, Anscombe's definition of intentional action is *mind-independent* to a significant extent. It defines intentional action as action which gives application to a specific form of linguistic interaction. In other words, it is not defined– at least not

explicitly – in mental terms. The trouble with maintaining that her definition is mind-independent only starts when we try to account for the fact that intentional action gives application to a specific form of linguistic interaction. For example, one might argue that a capacity to engage in this kind of linguistic interaction is a prerequisite for acting intentionally. So, for example, an agent's acting intentionally depends on his capacity to raise and answer the why-question about his own action. However, before jumping to such conclusions about intentional agency, we must note that Anscombe's definition of intentional action in terms of linguistic facts does not entail that intentional action *depends logically or explanatorily* on these facts. That intentional actions distinctively give application to a why-question might merely depend on (non-mental and non-linguistic) features of action which, if the why-question is raised, should be referred to in order to answer that why-question. For example, that intentional actions distinctively give application to a why-question might be a consequence of the non-linguistic fact that they, in contrast to other actions, are performed for a reason.

By means of which criterion can we distinguish between actions which give and actions which do not give application to a why-question for a reason? And how is it that we can distinguish rational action from mere action by means of this criterion?

Anscombe's own account of this criterion is somewhat confusing. On the one hand, she suggests that the why-question is given application by an action of an agent *iff* the agent can answer that question (Anscombe 2000 par 16). On the other hand, however, she allows that the why-question is given application by an action of an agent, e.g., an animal, who lacks a capacity to answer that question (Anscombe 2000 par 2, 47; Anscombe 1979). *Please note* that both characterizations are mind-independent, or at least not explicitly mind-dependent.

This allows for two Anscombian accounts of intentional action. Firstly, we could argue that Anscombe's theory involves two why-questions: a 'weak' one and a 'strong' one. An agent's intentional action gives application to the *strong* why-question *iff* he is capable of making explicit the reason for which he is acting. An agent's action gives application to the *weak* why-question, irrespective of whether or not he is capable of making explicit the reason for which he is acting. Correspondingly, there are weak and strong concepts of intentional action in Anscombe's account. On the strong concept, an action is intentional *iff* an agent is capable of making explicit the reason for which he is acting. On the weak concept, an agent's action is intentional *irrespective* of whether or not he is

capable of making explicit the reason for which he is acting, as long as the why-question applies.

Secondly, and alternatively, we could claim that there is only one why-question in Anscombe's account: the weak one (cf. Gustafsson). Correspondingly, there is only one concept of intentional action: the weak one. This would imply that, qua intentional action, there is no difference between actions of agents who can and cannot make explicit that they act for a reason. There would, qua intentional action, be no difference between an agent drinking coffee to stay alert and a bird landing on a twig to pick at some seeds. Certainly, this second reading leaves open the possibility that – as a matter of fact – some agents act intentionally *if* they are capable of making explicit the reason for which they act. However, the second reading would suggest that this possibility can be explained in at least two competing ways:

> (i)   his capacity to act for a reason *marks* his capacity to make explicit the reason for which he is acting; or
> (ii)   he has, independent of his capacity to act for a reason, a reliable capacity to make explicit the reason for which he is acting.

Which of these two readings offers the best prospects for an Anscombian account of intentional action? Anscombe herself offers no positive account of rational action to support the remarks she makes about non-linguistic animal action. However, she also suggests that acting for a reason depends on practical knowledge and practical reasoning. Should we hold that animals are capable of pre-linguistic practical knowledge or practical reasoning? Or, should we hold that Anscombe's remarks about animal action are somewhat confused? We could be tempted to argue that no matter what: intentional action depends on consciousness and reasoning, be it linguistic or pre-linguistic. Alternatively, we could argue that there is only a strong why-question and, correspondingly, only a strong concept of intentional action. Or, alternatively, we could argue that an Anscombian account of intentional action in terms of practical knowledge and practical reasoning remains untroubled by an account of the weak why-question and weak account of intentional action – if indeed such an account is possible.

To sidestep the weak intentional action account, or to stress that such action would still depend on practical knowledge or practical reasoning would, however, be somewhat question-begging. For Anscombe develops her account of intentional action in terms of

practical knowledge and practical reasoning, *only to explain the fact* that we human agent's non-observationally know what we are doing intentionally. We may of course argue that animals, which lack the linguistic capacities which allow us to distinguish intentional from unintentional action when it comes to human beings, may still have non-observational knowledge of their intentional action. However, the assumption that we or they act intentionally, does not warrant the assumption that intentional action depends on such knowledge. Hence, we are free to consider that intentional action – both strong and weak - is independent of practical knowledge and practical reasoning.

2.3 Non-observational Knowledge and Practical Reasoning?

Although Anscombe's definition of intentional action is mind-independent, her account of intentional actions of *linguistic* agents is not. Not only does she claim that such agents non-observationally know their intentional action, she also *explains* such knowledge in terms of an agent's practical reasoning on which his action depends.

Anscombe implies that an agent who is capable of making explicit the reason for which he is acting acts intentionally *only if* he non-observationally knows what he is doing (Anscombe 2000 par 16).[4] For example, an agent who is drinking decaffeinated coffee does so intentionally *only if* he non-observationally knows that he is drinking decaffeinated coffee. If we ask him, 'Why are you drinking decaffeinated coffee?' he could refuse that why-question application by saying, 'I thought I was drinking normal coffee!' Non-observational knowledge is a necessary but insufficient marker of intentional action. There are things other than actions which we know independently of observation. For example, one need not observe one's legs to know where they are. In addition, one need not observe oneself to know that one is in pain. One need not observe that one is drinking coffee to stay alert to know that one is doing so.

Anscombe explains intentional action on the assumption that it depends formally on implicit or explicit practical reasoning (Anscombe 2000 par 43; Moran 2004b; van Miltenburg 2012). Such practical reasoning, in turn, is part and parcel of an agent's practical knowledge which constitutes his non-observational knowledge of his

---

[4] Although often overlooked, Anscombe goes further and argues that such agents non-observationally know the *causes* of their intentional action if they act intentionally (Anscombe 2000 par 16). This leaves open (i) whether or not his knowledge of the reasons for which he is acting depends on his knowledge of these causes and (ii) whether this non-observational knowledge of causes is knowledge of efficient causes or of other forms of causes.

intentional action. For example, the agent who is drinking coffee to stay alert *non-observationally knows* that he is doing so because his doing it depends formally on his knowledge of what he is doing and why he is doing it.

Anscombe's account of strong intentional action yields two challenges to a mind-independent, generous account of rational action.

First challenge: explain that rational actions, though independent of a capacity to consciously act for a reason, give application to a why-question for reasons.

Second challenge: explain, without attributing to him a capacity to consciously act for a reason, that a (linguistic) agent acts for a reason *if* he (i) can make explicit the reason for which he acts and (ii) non-observationally knows what he does and why.

I will develop an argument in the remainder of this and the subsequent two chapters which would neutralize the first challenge, and I will implicitly address the second challenge in the final two chapters of this dissertation, where I develop an account of agential self-consciousness.

To neutralize the first challenge, let us return again to an observation which Anscombe makes about linguistic agents. A linguistic agent can deny that his action gives application to the specific why-question *on the ground that he did not know that he was doing it*. This might be taken to suggest that his non-observational knowledge of his action is essential for intentional action, at least if he is a linguistic agent.

No matter how appealing this suggestion is, it stands in some tension with the idea of a *weak* why-question and a *weak* concept of intentional action, which we would need in order to make sense of Anscombe's claim that non-linguistic animals can act intentionally. Firstly, given that it is far from obvious that animal agents non-observationally know what they do, it would be *surprising* if Anscombe proceeded as if non-observational knowledge was logically and explanatorily prior to their intentional action. Secondly, given that it is far from clear that non-linguistic agents are capable of practical reasoning, it would be equally surprising if Anscombe proceeded as if intentional action depended on practical knowledge and practical reasoning. If this is

so, then why should the linguistic agent's capacity to act for a reason differ from the non-linguistic agent's?

Therefore, we might ask Anscombe whether a non-linguistic agent should have non-observational knowledge of his rational action, and whether his acting rationally depends on a capacity for practical knowledge and practical reasoning. The problem is that Anscombe suggests that such agents have such knowledge. She claims that we ascribe intentions to a cat '[…] though the cat can utter no thoughts, and cannot give expression to any knowledge of its own action, or to any intentions either' (Anscombe 2000 par 16). This strongly suggests that she thinks that cats have at least some mental capacities on which their intentional action depends (cf. Sullivan 2009).

We might be fairly sceptical about attributing such mental-life to non-linguistic animals, and hence not opt to explain animal intentional action – if such action exists – via the assumption that animals have practical knowledge.

2.4 The First Challenge: Two Questions

Of course, this question and such scepticism are insufficient to neutralize the first challenge: explaining that rational agency is independent of a capacity to consciously act for a reason. It does not prove that we can or should replace Anscombe's full-blown account of intentional action with a mind-independent, generous account of rational agency. Anscombe's claim that animals act intentionally might be largely immaterial to her full-blown account of intentional action, i.e., if it yields problems for this latter account then she might easily drop her claim. In addition, inspired by her full-blown account, we might argue that non-linguistic agents might have a *rudimentary*, pre-conceptual, non-observational awareness of their intentional action, because their action depends formally on their pre-reflective practical reasoning or practical knowledge.

There is no need for me to insist that Anscombe definition of intentional action requires a weak account of action. On the contrary, for the purposes of developing a mind-independent, generous account of rational action, I thank her for defining intentional action in the way she does, and I will develop the thesis that her definition is compatible with a mind-independent account of rational action. To develop this thesis, I will have to complete two tasks. I will have to develop a mind-independent, generous account of acting in ways which give the specific why-question application. I will

explain such action in terms of motivational responsiveness to certain *facts* about that action which figure centre stage in reason-explanations (chapter 3). In addition, I will have to ensure that such facts and motivational responsiveness to them mark a capacity to act for a reason, independent of a capacity to consciously act for a reason (chapter 4).

2.5 Conclusion

My aim in this chapter was to set the stage for the sort of generous account of rational action which is needed in order to develop the dual capacity thesis. In the first *two* sections I distinguished mind-dependent from mind-independent accounts of rational agency. A mind-dependent account defines rational agency in terms of mental states/capacities or a subject regulating that action. A mind-independent account would not define rational action in terms of mental properties. A mind-independent account would offer better prospects for the dual capacity thesis. I introduced Anscombe's definition of intentional action as a mind-independent definition of rational action. In defining intentional action as action which gives application to a why-question for reasons, Anscombe appears to be neutral as to whether or not rational actions depend on mental capacities. In the third section I argued that Anscombe's own mind-dependent account of intentional action in terms of practical knowledge and practical reasoning stands in some tension with her claim that non-linguistic agents can perform actions which give application to the why-question. I argued that this tension allows us to raise the question of whether an agent needs a mind to act intentionally. In the short fourth section, I framed the task for the next two chapters: to explain, without appealing to conscious agency, the distinction between actions which give application to a why-question for reasons, and those which do not.

3 – Mind-independent Rational Action?

3.0 Introduction

Rational action, distinctively, gives application to a why-question for reasons. An answer to such a why-question for reasons, in turn, gives expression to a '... in order to ...'-conception of that action. This suggests that a rational action, *x*, distinctively happens because of what it brings about or because of mental representations of what it brings about. This characteristic of rational action might easily lead us to adopt the thesis that a rational action is determined by an agent's intention with it or that it is caused by an agent's beliefs and desires. However, because intentions, beliefs and desires are mental states, a generous account must explain the '... in order to ...' component of rational action differently.

In the present chapter I will develop the thesis that an '... in order to ...' conception of an action applies to an action if that action happens because an agent is motivationally responsive to a reason. This chapter is divided into four parts. In the *first* part I will develop the thesis that an action, *x*, gives application to a *why*-question for reasons *if x*

happens because *x* is a factor for something else (*s*). In the *second* part I will interpret this thesis in terms of the assumption that *reasons for action* are external to the mind and, consequently, that *x*-ing *for a reason* amounts to *x*-ing because one is *motivationally responsive* to the fact that *x* is a factor for *s*. In the *third* part I will develop the view that such motivational responsiveness is independent of mental capacities. In the *fourth* section I will argue that the generous, mind-independent account is not overly generous, but it is robust enough to distinguish rational actions from mere happenings for which we give ordinary causal explanations, on the one hand, and merely functional events and reflexes for which we might give teleological explanations, on the other hand.

3.1 The Why-question and Facts which Answer It

Which feature of a rational action gives application to an '... in order to ...'-concept of that action? How are we to analyse expressions such as 'he is drinking coffee *in order to* stay alert'? In the literature, we find, broadly, two answers. Some philosophers would suggest that such concepts apply to an action *if* that action is *determined* by an agent, i.e., not merely by his mental states. Some argue that such determination is non-causal, but is, say, teleological, others, that it is causal in a special way, say agent-causal (cf. Anscombe 2000; Hornsby 2004; Gustafson 2007; Chisholm 1976; Velleman 1992). Others would argue that '... in order to ...'-concepts apply to an action *if* that action is *caused* by mental states, such as pro-attitudes, beliefs, intentions or plans which *justify* that action (cf. Davidson 2006a; Bratman 1987; Mele 2000). This latter group of philosophers favour a 'causal theory of action'.

I will sidestep a choice between these two approaches for the present moment because, in the context of this chapter, I must avoid both. This is because both approaches are mind-dependent accounts of rational action. Causal theorists would claim that we can apply an '... in order to ...'-concept to a rational action only because it is caused by *mental states which justify that action*. Those who oppose this causal theory of action do so mainly to emphasize that agents are involved in their action *over and above the causal effects of their mental states*.

Therefore, if we are to develop a mind-independent account of rational action, we must go beyond these two approaches to action. Yet we must still answer the following question: which feature of rational action gives application to an '... in order to ...'-concept of that action? In order to develop a mind-independent account of action which

answers this question, I will explore Taylor's analysis of 'teleological explanation', on the one hand, and 'behaviour explanation', on the other hand (Taylor 1964; for comments see Noble 1967; Taylor 1967).

Taylor emphasizes that behaviour belongs to a class of happenings which, in contrast to *mere events* which are to be explained by reference to other events which cause them, are to be explained by reference to what they bring about. That is to say, behaviour requires, according to Taylor, 'teleological explanation'. He leaves open whether the class of behaviour explanation exhausts the class of teleological explanation (Wright 1973; Wright 1972; Stout 1996). Notwithstanding this, in his view, behaviour requires a *specific* sort of teleological explanation, i.e., what I call 'behaviour explanation'.

According to Taylor, a *teleological explanation* of *x* explains *x* on the ground that *x* is *necessary for something – s – to occur*. On this analysis, a teleological explanation *does not* explain *x* on the ground of *s*. Instead, a teleological explanation of *x* explains *x* on the ground that *x* is *necessary* for *s* to occur. That is to say, a teleological explanation of *x* may successfully explain *x* on the ground that *x* is necessary for *s* to occur, even if *x* does not in fact bring about *s*.

As mentioned before, Taylor emphasizes that behaviour distinctively requires explanation by purpose (cf. Taylor 1964, 27). An explanation by purpose explains behaviour *x* as something 'directed' by an agent to realize *s*. On Taylor's view, such directing depends on the agent's basic goals. To explain the possibility of such guidance, Taylor attributes desires and intentions to the agent– which he emphasizes might involve but are independent of reflective capacities (Taylor 1964 in particular the second and third chapter).

In the context of this chapter, I am interested in Taylor's claim that behaviour explanation is a species of teleological explanation. However, I think that there is a gap to be explored between his analysis of mere teleological explanation, on the one hand, and explanation by purpose, on the other hand. More specifically, it is unclear whether, and if so how, we must analyse behaviour explanations in terms of his account of teleological explanation. On my reading of his account, a teleological explanation explains *x*, on the ground that *x* is *necessary* for an *event/state – s –* to occur. Taylor's explanation by purpose, in contrast, explains a behaviour, *x*, on the ground that an agent directs it to realize his goal – s. Is behaviour explanation really teleological? If so,

then what happened to the *necessity* because *of* which teleological events happen? What made *s* depend on a goal of an agent which underlies an agent's desire or intention to bring about *s*?

As a first attempt at recasting Taylor's account of behaviour explanation in terms of his account of teleological explanation, we might argue that behaviour explanation of *x* is teleological, insofar as it explains *x* on the ground that *x* is *necessary* for *s* to occur, where *s* is a goal of an agent. In contrast to mere teleological explanation, behaviour explanation explains behaviour on the assumption that an agent is motivated to do *x* by the fact that *x* is necessary for his *goal - s*. Such motivation, on Taylor's view, involves an agent directing his action on the ground of his desire for *s*. Whether or not an agent causes *x* on the ground that *x* is necessary for *s*, depends on what I will call an agent's 'motivational responsiveness' to the fact that *x* is necessary for *s*.

Taylor's idea that an agent can act only because he has a desire/intention would imply that such *motivational responsiveness* depends on a conscious capacity for intentional action. This is, obviously, one way to make sense of the *motivational responsiveness* assumption in behaviour explanation. However, such an interpretation of this idea of motivational responsiveness would, although generous, favour a mind-dependent account of agency.

We can develop an analysis of behaviour explanation, and in particular of motivational responsiveness, which allows us to develop a mind-independent account of rational action from Taylor's account of teleological explanation. To do so, let us explore the logical space which the notion of motivational responsiveness provides for. Some space does exist, because Taylor's account of a teleological explanation of *x* explains a teleological process, *x*, in terms of the fact that *x* is necessary for *s*, independently of *mental capacities*. In particular, such processes do not depend on *s* being a basic goal at all, let alone desired. Therefore it should be an open question as to whether or not an explanation of *x*, on the ground of motivational responsiveness to the fact that *x* is necessary for *s*, depends on mental capacities at all.

In the remainder of this section, I will scrutinize several interpretations of Taylor's account of teleological explanation and introduce the sort of reading which would align with a mind-independent, generous account of rational action. This will form the

background against which I will introduce the idea that *reasons for action* and *acting for reasons* are mind-independent.

Quite recently, Rowland Stout has developed an account of action based on Taylor's account of teleological explanation (Stout 1996). Stout argues that an action is something that happens because it *should*. In his view, whether or not an event is an action depends on whether or not it happens *because* it should. Whether or not an action should happen, in turn, depends, according to Stout, on the factual conditions in which the agent is acting. On Stout's conception of it, a capacity to act would be best understood as a capacity to find out whether or not, given the facts, one *should* do *x*. An agent with such a capacity must be conceived as someone who (i) has a repertoire of interrelated questions which allows him to determine what he should do in light of the facts and (ii) a capacity to answer this question in light of the facts, so as to do what he should (Stout 2006).

For example, an agent who is trying to stay alert will have a repertoire of questions about the facts available which help him determine what he should do. He may ask whether he has time to sleep *or* whether there is some coffee in the jar. He may find that his schedule is too tight, so that he should either cancel an appointment or not have any sleep. In addition he may find that there is no coffee in the coffee jar. Faced with this, he may ask himself whether he should make some coffee himself or get some from the café just outside his office. He may know that the answer to the first question depends on whether there is any coffee in the cupboard and that the answer to the second question depends on whether there is any money in his wallet. He may find that there is no money in his wallet, but that there is coffee in the cupboard. Facing these facts, he will decide – given the facts – that he should make himself some coffee.

Before relating this aspect of Stout's account of intentional action to Taylor's account of teleological explanation, it is important to understand it in light of another feature of Stout's account of action (Stout 2005). On Stout's account, actions are *not* single events *or* causal sequences of such events which are caused by motives, but rather a series of events which (i.e., the individual events) are caused because they should be caused. On his model, for example, whether or not an agent drinks coffee intentionally depends on whether or not the agent causes the pouring of the coffee, the lifting of his cup and his taking a sip, etc., because – given the facts – he should cause these events to, say, stay alert. On Stout's account, whether or not an agent is performing a specific action

depends – I would say – on whether or not he is, throughout the action, motivationally responsive to the facts: asking himself during the whole sequence of events which he causes, whether or not he should continue, adjust or stop the series of events which he caused.

In this respect, Stout's criteria for calling an action 'intentional' differ from Davidson's. Imagine someone who is drinking coffee to stay alert. While drinking the coffee, the cup drops from his hands on to his trousers. The hot coffee burns his leg in ways which increase his alertness. Did he increase his alertness intentionally? On Davidson's account, it seems that the agent is drinking coffee because he desired to stay alert and believed in drinking coffee as a means to stay alert. The agent's mental states are not only causes of his drinking coffee, but also of his dropping the cup and becoming alert. What is more, they justify what they cause. Hence, Davidson would imply that our agent intentionally increased his alertness. Of course, this is a well-discussed, counter-intuitive implication of Davidson's theory. The core remedy to it seems to be to argue that an intentional action is caused *in the right kind of* way by mental states which justify that action. Issues like these are discussed by philosophers under the label 'wayward causation' or 'deviant causal chains' (cf. Mele 1987; Schlosser 2007).

I need not discuss the problem of causal deviance here. I merely mention it to clarify the second distinctive feature of Stout's account of rational action: that a rational action is typically part of a series of events, each of which is caused by an agent because it should be. Stout's account would imply that the agent did not intentionally increase his alertness, because his dropping the cup was not part of the series of events which (i.e. the events) he caused because they should have been caused. Although dropping the cup was – partially – the effect of an event (e.g., raising the cup) which he caused because it should have been caused, the dropping itself was not caused because it should have been.

Stout's account – unlike Davidson's – implies that an agent is involved in his action not because his mental states caused it, but because he is continuously concerned with what he is causing, asking whether or not he should do something additional, interfere with it or compensate for something so as to complete a course of action. This aspect of Stout's account neatly displays what I called the 'diligence' which marks action above mere activity.

Hence, Stout can distinguish an action from (i) a mere happening, (ii) a mere sequence of mere happenings, (iii) a mere activity and (iv) a mechanistically controlled series of events. An example of the first event would be a raindrop falling from the sky. Something that fits the description of the second would be a raindrop falling from the sky on a leaf and dripping from that leaf into a pool. An example of the third might be Velleman's president saying, 'Hereby, I declare the meeting closed.' An example of the fourth would be the series of pulses caused by a thermostat which cause a heater to produce heat or to stop it from producing heat. All these fall short of being action, because they are caused independently of whether they *should* happen.

In terms of developing a mind-independent, generous account of rational action, I am largely sympathetic to Stout's revival of Taylor's account of teleology and, in particular, his account of action in terms of it. More specifically, I am sympathetic to his idea that a teleological explanation of *x* explains in terms of *facts*. In consequence, I am sympathetic to the idea that actions are to be explained in terms of a specific class of facts. Moreover, I adopt, for the purposes of spelling out a generous account, Stout's thesis that an agent acts only if he is continuously involved in what he is *doing* causing a series of events.

Nevertheless, Stout's account of action is much too mind-dependent to develop a generous account of rational action from it. Part of this mind-dependence is reflected in Stout's emphasis on actions being things that happen because they *should*. More particularly, it is reflected in his account of the capacities which an agent needs to derive from the facts what he should do. On Stout's view, an action explanation cites a fact which entails a *practical requirement* to perform an action. Stout ascribes to an agent a capacity to raise and answer questions about facts in order to infer what he *should* do. In this respect, Stout's account of teleology is much stronger than Taylor's account of behaviour explanation. It is not to be assumed that motivational responsiveness to the fact that *x* is necessary for *s* depends on capacities to raise and answer questions about facts so as to determine whether *x* is necessary for *s*, let alone to determine whether, given the facts, *x* should be done.

Let us investigate more closely the terms in which Taylor explains actions and, more specifically, whether a teleological explanation of *x* should explain *x* by the fact that *x* is necessary for *s*. Here it is necessary to discuss Larry Wright's criticism of Taylor's

account of behaviour explanation, and teleological explanation in general (Wright 1972; Wright 1973). Wright would summarize Taylor's account of teleology as follows:

> x occurs for the sake of s means:
>> (i)   x is necessary (required) for s to obtain;
>> (ii)   x's being necessary for s is sufficient for x to occur.

Wright argues that we can imagine a variety of scenarios in which an agent does *not x*, even though *x* is necessary for *s* to occur. He might be locked away, be doing other things or be too exhausted, etc. Hence, *x*'s being necessary for *s* is *not* sufficient to explain *x*. We could – as Taylor does – explain *x*'s not occurring in terms of 'factors interfering' with the teleological process. This would imply that *x* being necessary for *s* is – in the absence of interfering factors – sufficient for *x* to occur. However, Wright stresses that this analysis of behaviour explanation would be problematic. It would entail a problematically generous concept of interfering factors. For if being locked away, doing other things and being exhausted would all count as interfering factors, then on Taylor's analysis it would be impossible to distinguish causing *x because x is* necessary for *s*, on the one hand, from *accidentally* causing *x* when *x* is necessary for *s*, on the other hand. For example, a bird may accidentally collide with a cat's claws, even though the cat merely exposed his claws because he was washing himself. If – in this situation – the cat has a desire to catch birds, and his exposing his claws was necessary for him to catch one, then was his exposing his claws part of a teleological process by which he cancelled an interfering factor (i.e., washing himself)?

Wright argues that Taylor's account has other problems as well. It would not allow us to distinguish an *agent's doing y from his doing x* if (i) while *x*-ing, he pauses from doing *x* and starts doing y, when (ii) what he did in terms of *x*-ing was what was necessary for him to be able to do *y*. For example, imagine that a cat walking to its bowl notices a pool of water from which he drinks. Was the cat – before he noticed the pool – really walking in order to get some food or was he walking in order to drink something from that pool?

In addition, Wright claims that Taylor's account cannot distinguish intentional action from merely functional processes. For example, although it is necessary for the cat to sleep for him to perform any activity at all, there is an important sense in which the cat

sleeps because sleep is necessary in order to engage in activity, even though the cat does not sleep with an intention of doing these things.

To circumvent these problems, Wright proposes to revise Taylor's account of teleology in the following way (Wright 1972, 211):

> x occurs for the sake of s means:
> (i)   x tends to bring about s;
> (ii)   x occurs because […] x tends to bring about s.

From this, he infers an account of behaviour explanation:

> Agent A does x for the sake of s means:
> (i)   x tends to bring about s;
> (ii)   x occurs because […] x tends to bring about s.

'Because' in the second clause of both analyses is used in the ordinary sense of denoting a causal relation. Wright therewith emphasizes that a cause may occur without the effect. In addition, Wright claims that a behavioural or a teleological process, *x*, is caused by *x*'s *tendency* to bring about *s.* Hence, his account implies two things: (i) that even though *x* tends to bring about *s*, *x* might – sometimes – not occur when *x* tends to bring about *s*, *and* (ii) that when *x* occurs, it might sometimes not bring about *s*.

Wright's analysis implies that *x* is a teleological process *only* if *x* is caused – in the ordinary sense of causation – by *x*'s tendency to bring about *s*. As a result, Wright circumvents the problems which he attributes to Taylor's account.

We now have *two* notions in terms of which we could interpret Taylor's account of action explanation as an account of teleological explanation. On the one hand, we have Stout's interpretation that a behaviour explanation of *x* explains *x* on the ground that it *should* happen. On the other hand, we have Wright's claim that a teleological explanation of *x* explains it not on the ground that that behaviour is *necessary* but on the ground of its tendency to bring about *s*. Stout's account, as I argued before, is much too mind-dependent. However, Wright's account of behaviour explanation has

problems of its own. More specifically, it fails to analyse *two* distinctive features that intentional action has, in contrast to other functional features, processes or events. That is, it fails to analyse the role of an agent in his action. What is more, Wright's account seems to suggest that action explanations should explain x either in terms of previous x's causing s or as an instance of an action, type X, tokens of which are explained by the fact that generally they cause instances of s. Although we might seek to give such explanations for actions, doing so would not allow us to distinguish actions from merely functional processes (cf. McLaughlin 2001). To distinguish actions from merely functional processes, we might turn to the fact that we normally explain an intentional action - x_1 (but not functional processes) also and predominantly by reference to the fact that this particular x_1 *causes or constitutes* an instance of S (cf. Stout 1996).

The issue at stake, however, is *neither* who provides the best interpretation of Taylor's account *nor* who offers the right theory of action explanation. I discussed Stout's and Taylor's accounts mainly to find some additional resources for developing a mind-independent account of rational action.

As mentioned above, Taylor emphasizes that a behaviour explanation of x is a *specific sort of teleological* explanation. For example, a behaviour explanation of x would explain x on the ground that an *agent does x* because x is necessary for s to occur. Taylor explains this in terms of the desires which an agent has and which motivate him to do x because x is necessary for s. And this, I think, saves him from Wright's criticism. Taylor's analysis of behaviour explanation in terms of motivational responsiveness is open to two readings. On the first reading, the fact that x is necessary for s to occur sufficiently explains that an agent does x because he is motivationally responsive to it. This reading would entail, for example, that whenever drinking coffee is necessary to stay alert, there is an agent who is motivated to drink coffee in order to stay alert. This would be absurd, of course, given that an agent might not be motivated to drink coffee, even if doing so is necessary for him to stay alert. On a second reading, the fact that x is necessary for s to occur is necessary, but it is insufficient for an agent to do x. That is, although drinking coffee is necessary to stay alert, there might not be an agent who is drinking coffee. That is to say, there might not be an agent who is motivationally responsive to the fact that drinking coffee is necessary to stay alert.

In contrast to the first reading, the second reading of behaviour explanation would circumvent Wright's criticism of Taylor, but only because it assumes that the fact that x

is necessary for *s* is *never sufficient* for *x's* occurrence. It would, for example, allow us to argue that, despite the fact that drinking coffee is necessary to stay alert, our agent is not drinking coffee, *because he is not motivationally responsive to that fact*. An analysis of behaviour explanation would then look something like the following:

> 'Agent A does *x* in order to *s*' means:
>   (i)   *x* is necessary to bring about *s*;
>   (ii)   *x* occurs because A is motivationally responsive to the fact that *x* is necessary to bring about *s*.

In contrast to Wright's analysis, this analysis pays explicit attention to the role of an agent in causing his behaviour.

As a consequence, this second interpretation – indeed, revision (cf. Taylor 1964, 27) – of Taylor's account of behaviour explanation in terms of motivational responsiveness would allow us to raise the question of whether we should agree with Taylor that agents need mental capacities to do *x* because of the fact that *x* is necessary for *s*. Must an agent – as Taylor has it – have desires, let alone a mental capacity to act in light of his desires? Must he – as Stout has it – have capacities to infer from the facts whether or not he should do something? Why should he not merely have capacities by which he is motivationally responsive to the fact that doing *x* is a factor for *s*?

Wright's account seems to go in this latter direction. It leaves open by which capacities an agent manages to act purposively, and does away with the suggestion that an agent does *x* intentionally only if he does *x because x* is necessary – be that factually or normatively.[5] In this respect, Wright's account does more justice to the facts: facts do not dictate what is necessary, and they typically leave ample room for alternative ways of realizing *s*. For example, to stay alert one may drink coffee, drip hot coffee on one's trousers, hammer one's toe with one's cup, and who knows what else. The facts do not dictate what should be done. Apparently, we could have an account according to which an action, *x*, would already be rational – i.e., give application to a why-question for

---

[5] It might be that to determine whether or not an agent acts for a purpose, we need an environmental design in which the difference between behaving mechanically and purposively is the difference between not doing and doing what should be done *or* what it is necessary to do for the realization of state *s*. This, however, does not prove that the agent's doing what should be done or what is necessary *depends on his being motivationally responsive to its being required or necessary.* Instead, he might merely be doing *x* when it is required or necessary for *s* to occur *because* he is motivationally responsive to the fact that *x* tends to bring about *s*.

reasons – if an agent does *x* because *x* 'tends to' bring about something (*s*). We would then get the following analysis.

> 'Agent A does *x* in order to *s*' means:
> (i)   x tends to bring about *s;*
> (ii)   *x* occurs because A is motivationally responsive to the fact that *x* tends to bring about *s*.

Therefore, Wright's account, even though his account is flawed in several other respects, has important features which are absent from Taylor's and Stout's.

In terms of developing a generous account of rational action, I am somehow caught between Taylor's notion of 'necessity' and Wright's notion of 'tends to'. For the reasons cited above, I think that rational actions do not happen because they are necessary or required. On the other hand, I would agree with Stout that the notion of 'tends to' does not allow us to explain a concrete action, x_1, on the ground of facts which obtain here and now. Our coffee-drinker does drink coffee because – in the circumstances in which he drinks it – his drinking coffee is a factor for alertness to occur. The above analysis does not capture this aspect, at least not on both of the following two readings of it. According to the first reading, an agent performs x_1 because tokens of type X at previous times caused tokens of type S. For example, our agent is drinking coffee because, previously, instances of coffee-drinking caused instances of alertness. Such an explanation would be a sort of causal explanation of his present coffee-drinking.

According to the second reading, an agent performs x_1, a token of X which (i.e., the type) exists because generally tokens of *X* cause tokens of *S*. For example, my drinking coffee is a token of a type of action, i.e., coffee-drinking, which is explained by the fact that its tokens generally cause instances of alertness. This explanation explains x_1 by reference to facts about the type of action X rather than in terms of the particular instance x_1 which we seek to explain here.

There is, however, a third revisionary reading of Wright's account, which would address Stout's worries independently of giving in to the idea that actions happen because they should or because they are necessary. According to this third reading, an agent performs x_1 because x_1 is a *factor* for bringing about *s*. For example, I am drinking coffee because my current coffee-drinking is a factor for doing that which brings about

increased alertness. This third explanation differs from the earlier two in an important way. It explains x_1 not merely on the ground of historical instances of X causing instances of S, nor on the ground that tokens of action type X generally cause instances of S, but primarily on the ground of the fact that this particular token, x_1, is a factor for realizing s. Moreover, the 'factor of' reading – unlike the above two 'tends to' readings – does not explain x_1 on the assumption that x_1 is necessary or (tends to be) sufficient for s.

> 'Agent A does x_1 in order to s' means:
> (i) x_1 is a factor for s to occur;
> (ii) x_1 occurs because A is motivationally responsive to the fact that x_1 is a factor for s to occur.

On this analysis, if x_1 is a factor for s, and x_1 happens because an agent is motivationally responsive to that fact, then x_1 1 might still occur without s. In addition, if x_1 is a factor for s, then this analysis leaves open whether or not there is an agent who does x_1 because of the fact that x_1 is a factor for s. In what follows, I argue that this analysis of action explanation allows us to develop a mind-independent, generous account of rational action.

Let me finish this section with two remarks about what it is for an action to be a 'factor for s to occur'. The variable s might refer to anything whatsoever, as long as it can partly or wholly be caused or constituted by x. For example, s might refer to a particular state of affairs such as alertness, an event such as becoming alert or a fact such as that one is alert. However, it might also refer to facts about relations, such as that s refers to something which is generally a factor for something else. For example, an agent's drinking coffee to stay alert might best be analysed in terms of the fact that he is doing what is doing because it is a factor for that which is generally a factor for staying alert. In addition, a factor for s might, but need not, be a cause of s. Although it might be a (eventually insufficient and unnecessary) causal factor for s, it might not, and instead it might be a constitutive factor for s (cf. Taylor 1964, 27; Balaban 1990). Therefore, when I use 'factor' I mean *causal* or *constitutive* factor; what it is a factor of might be whatever can be caused or constituted at all.

Perhaps it should be mentioned explicitly that on this analysis of action explanation, s can refer to x. So, for example, an agent can drink coffee, because drinking coffee is a

factor for drinking coffee. Drinking coffee on these grounds would be rational, and distinct from, say, accidentally drinking it, being mechanistically forced to drink it or drinking coffee caused by other processes.

3.2 Reasons and Motivational Responsiveness to Reasons

It is clear that a teleological account of behaviour explanation rationalizes that behaviour. A behaviour explanation of x_1 *rationalizes* x_1 insofar as it explains x_1 (partly) by the *fact* that x_1 is a factor for s to occur. However, that fact, while it rationalizes x_1, is *insufficient* to explain the occurrence of x_1. That is to say, a behaviour explanation explains x_1 on the ground of an agent's *motivational responsiveness* to the fact which can be cited as rationalizing x_1. A behaviour explanation of x_1 explains it in terms of two relatively independent types of grounds: *motivational responsiveness* and *facts*. It is only if an agent's motivational responsiveness has particular features that he will do x_1 if x_1 is a factor for s. In other words, the fact alone that x_1 is a factor for s is insufficient to explain the occurrence of x_1. By the same token, an agent who does x_1 when x_1 is a factor for s might not be motivationally responsive to the facts which rationalize x_1 at all.

A similar distinction between grounds which rationalize an action and grounds which explain that action's occurrence has been stressed by Alvarez, who distinguishes between facts which *rationalize* and facts which merely *explain* (Alvarez 2010). According to her, facts might explain an action independently of rationalizing it. However, some facts which rationalize an action might be part of a causal explanation of that action. For example, we may explain that an agent drinks decaffeinated coffee in terms of the fact that he believes that his drinking coffee is a means to stay alert. This fact, however, does not rationalize his drinking decaffeinated coffee. However, we may explain that an agent drinks caffeinated coffee in terms of the fact that drinking coffee is a means to stay alert. This fact does *rationalize* and *explain* his drinking coffee if – in my terms – we assume that an agent is motivationally responsive to it. Alvarez suggests, however, that for such rationalizing facts to explain an action of an agent, we must assume that he has mental capacities by which he is motivated by these facts. But she also stresses that a rational action of an agent is genuinely determined by reasons which are external to his mind. For the purposes of developing a mind-independent, generous account of rational action, I am sympathetic to Alvarez's externalist distinction between facts which rationalize and facts which merely explain, i.e., her claim that the facts which rationalize are external to the mind. Moreover, I am

sympathetic to her assumption that we can explain an action in terms of such a rationalizing fact if – in addition to it – we assume a motivational system which causes that action on the ground of that fact. If I am to develop a mind-independent account of rational action, I must avoid, however, her suggestion that a motivational system thus responsive to facts involves mental capacities.

As a corollary, the distinction between facts which rationalize and facts which explain sheds new light on an old controversy between non-causal and causal theories of action. On this distinction, reasons are distinct from the motives that cause an agent to act for a reason. Reasons are insufficient to explain an action, and motivational responsiveness does not rationalize action. Instead, both causal and non-causal theorists are partly right: it is only if we assume that an agent is motivationally responsive to reasons that the reasons for which he acts indicate the motivational states in terms of which we must explain the occurrence of his action. By the same token, it is only on the assumption that he is motivationally responsive to reasons that his motivational states indicate the facts which rationalize his action.

Finally, it is quite important to understand that the 'factor for' account of behaviour explanation is immune to the intuitive criticism that teleological explanations assign causal or otherwise determining powers to goals. Bittner, for example, objects that a teleological explanation of x_1 explains x_1 on the ground of a state, s, which would exist only as the effects of x_1 (Bittner 2001). Such an explanation would be incoherent, in his view, because to understand it we would need to explain x as the effect of state s, which is caused by x_1. I agree with Bittner that we should not explain a rational action on the ground of things which do not obtain prior to or at the moment of action. In what follows, I will argue that the 'factor of' account of behaviour explanation is compatible with Bittner's demand. I will argue that the facts, in terms of which this account explains rational action, obtain at the moment of x_1 being carried out, quite independently of whether or not s obtains or will ever obtain.

Consequently, the 'factor of' account of behaviour explanation might not fit the standard picture, or caricature, of teleological explanation. That is to say, it does not assume that we must assign goals to an agent prior to conceptualizing and explaining his action in terms of reasons. It is, however, quite immaterial in the present context whether or not this account is a causal theory or a teleological theory of action.

3.3 Motivational Responsiveness

How are we to account for motivational responsiveness to the fact that x_1 is a factor for s, which explains that an agent is doing x_1? Does it depend on an agent's motivating a mental representation of that fact, such as a desire for s *or* an intention to s? As mentioned before, it is commonly assumed by philosophers that rational action depends on such mental states. Even philosophers who assert that reasons are external to the mind seem to imply that acting for 'externalities' depends on consciousness of them.[6] Certainly, many would endorse a generous concept of consciousness, emphasizing that these mental capacities required for acting for a reason might, but need not be, reflective. But I must still avoid these models of motivation by reasons if I am to develop a mind-independent, generous account of rational action.

So I must avoid this common philosophical answer, on the one hand, but on the other hand not collapse into traditional behaviourist, stimulus response models of action explanation. Let me repeat, I should contradict neither the possibility of rudimentary or developed forms of conscious rational agency nor mere stimulus-response behaviour. However, I must not generally assume that acting for a reason as such depends on mental capacities. In addition, I should not generally deny that apparent rational action can be understood in terms of stimulus-response models.

Whether or not motivational responsiveness depends, factually, on mental capacities should be an open question. However, I must develop the thesis that motivational responsiveness to the fact that x is a factor for s can cause x *independently of mental capacities.*

I would say that an agent is doing x_1 because he is motivationally responsive to the fact that x_1 is a factor for s if his doing x_1 has the following two phases:

(i)   he *initiates* x_1, independently of the fact of whether or not x_1 is a factor for s;

(ii) he continues doing x_1, depending on whether – factually – his doing x_1 is a factor for s.

---

[6] Bittner might form an exception to this, but his account of rational action as an agent's *response* to states or events does not provide for an explanation of the capacities which enable an agent to respond to such states or events.

On this analysis, the initiation of x_1 would not depend on (motivational responsiveness to) the fact that x_1 is a factor for s. It might merely be an agent's (conditioned) reflex to other facts, events or states which obtain prior to his doing x_1. However, for x_1 to be a rational action, it *should* be initiated by a motivational system which would be responsive to the fact that x_1 is a factor for s as soon as an agent realizes such a fact by doing x_1. That is to say, once he initiated x_1, an agent's x_1-ing progressively should depend on whether or not his x_1-ing causes or constitutes s.

The distinction between motives for initiating x_1 *and* motives for progressively doing x_1 depends on the assumption that doing x_1 takes time. Only because x_1 takes time can we deny that the reasons for which an agent is doing x_1 *should be either states which obtain prior to his doing it or his goal in doing it.* This distinction allows us to deny this, because it allows us to deny the assumption that the motives which cause rational action are uniform in kind. This assumption would require us to explain how it is that an agent can initiate his action, x_1, on the ground of the fact that x_1 is a factor for s. To explain this, we would need to ascribe to an agent a capacity to infer from conditions which obtain prior to his doing x_1 that, for example, doing x_1 is a factor for s. To explain how such inferences motivate an agent to do x_1, we would have to assume that he has a motivating concept of state s, based on which he does x_1 because x_1 is a factor for s. On such a model, although when he initiates x_1 there is no fact that x_1 is a factor for s, a rational agent would need to have a concept of this fact, even a justified concept. We might be tempted to say that such agents have a capacity to initiate x_1 *because x_1 is a factor for s.* However, strictly speaking, this would be wrong. They would initiate actions because they predicted that x_1-ing is a factor for s. No matter how reliable these predictions, they would not initiate x_1 because x_1 is a factor for s. Such agents would initiate x_1 because they predict that x_1 is a factor for s, *and would do x_1 because x_1 is a factor for s* only to the extent that their progressively doing x_1 depends on the fact that their doing x_1 is a factor for s.

I am happy to accept that such predicting agents are possible, as well as that they would be far more successful than knee-jerk agency *or* agents who are merely capable of continuing, changing or stopping to do what they are doing depending on that their doing it is a factor for s. However, for the purposes of developing a generous account of rational action, we need a concept of rational agency which allows us to deny that such predicting agents define the class of rational agents. More specifically, although knee-jerk agency is not rational, motivational responsiveness to the fact that x_1 is a factor for

s is independent of a capacity to anticipate this fact prior to initiating x_1. A mindless agent might not be able to initiate x_1 on the ground of the fact that x_1 is a factor for s, yet he might be able to progressively x_1 or *complete x_1*, depending on whether x_1 is a factor for s. Which motive initiates x_1 is immaterial to the question of whether or not he does x_1 for a reason, i.e., it might merely be an innate or acquired reflex to initiate certain actions. What matters in terms of whether he is doing x_1 because x_1 is a factor for s is whether an agent's progressively doing x_1 depends on the fact that his doing x_1 is a factor for s.

This analysis of rational agency assumes that reasons for action obtain at the moment of action. However, one might argue, the agent who is drinking coffee in order to stay alert would apparently become alert only when he had finished drinking it. Hence, his drinking coffee to stay alert appears to be independent of the fact that his drinking coffee is a factor for staying alert. So, one might think, the above analysis of rational actions fails for acts like drinking coffee in order to stay alert. This might not be so, however, for two reasons. First of all, on the above analysis whether or not an agent was drinking coffee to stay alert depends ultimately on what he does when lifting the cup and sipping from it, etc. is not followed by increased alertness. If he does nothing in response to the fact that his drinking coffee did not increase his alertness, then *either* he was not drinking coffee in order to stay alert *or* his motives changed in the interval between his progressively drinking coffee and the state in which his alertness is not increased. Secondly, and relatedly, an agent's drinking coffee in order to stay alert might merely mean that he is causing events (e.g., raising the cup and sipping from it) because these are factors for the sort of thing (i.e., drinking coffee) which is generally a factor for increased alertness.

What does it take for an agent to be progressively doing x_1 because his doing x_1 is a factor for s? He does not necessarily need a mental motivational system. He merely needs a motivational system in which:

      (i)   a motivational state, m_0, causes a motivational state, m_1, which, in turn, causes x_1; and

      (ii)   a motivational state, m_2, caused by x_1's being a factor for s, which maintains m_1.

The key to accepting that motivational responsiveness to reasons can be mind-independent lies in appreciating that this motivational system need not be mental. Still, however, an agent endowed with such a motivational system can be said to be x_1-ing progressively because doing x_1 is a factor for s. He need not have any consciousness of the relation between his doing x_1 and s, but his progressively doing x_1 should depend on a fact which he brings about by initiating x_1, i.e., that doing x_1 is a factor for s. His doing x_1 because he is motivationally responsive to the fact that x_1 is a factor for s might – but need not – involve a mental capacity.

This account of action does not collapse into stimulus response behaviourism, as it does not explain an action as a response triggered by sensory stimuli.

Before concluding this chapter, let me clarify the difference between a behaviour explanation in terms of motivational responsiveness to the fact that x_1 is a factor for s, on the one hand, and other forms of teleological explanation, on the other hand.

3.4 Action Explanation vs Explanation by Function

In this final section I will clarify the difference between the model of action explanation introduced above and an explanation of merely *functional events or reflexes*. Functional entities are defined in terms of what they bring about. So, for example, the heart's function is to pump blood. The function of breeding is to generate offspring. The function of the pupil reflex is to prevent eye damage resulting from too great a light intensity. If function ascriptions explain, then apparently ascribing a function to x amounts to explaining x in terms of what x brings about. If so, how are we to distinguish explanation by function from action explanation? Or are we to hold that we are to explain action in the same way as we explain functions or functional reflexes?

Some philosophers deny that function ascriptions explain at all (cf. Mitchell 1993; Ehring 1985; Minton 1975). Some say, for example, that ascribing a function to x amounts to characterizing x in terms of its causal role in a system (Couch 2011). For example, saying that the heart's function is to pump blood amounts to saying that the heart is the cause of blood circulation in, say, a functioning human body.

Others argue, however, that we do explain x if we ascribe a function to it (cf. McLaughlin 2001; Wright 1973). That is, we give an explanation of x in terms of what x brings about. There is disagreement, however, about whether such explanations reduce to

ordinary causal ones *or* are teleological. Some teleological explanations are relatively unproblematic, e.g., the ones we give for events or entities which – supposedly – depend on conscious agency. These include bodily movements, such as nodding one's head, or artefacts such as scissors or churches. To explain such events or artefacts, we can postulate a conscious agent who has a purpose for them. The problematic cases of teleological explanation would be those for which we cannot postulate conscious agency. How can a mind-independent feature, process or reflex be explained in terms of what it brings about?

One familiar answer to this question is broadly evolutionary: because *x* characteristically brings about *s*, it is the cause of itself. We could argue that our heart has a function to pump blood, because its pumping blood causes its own reproduction. That is, our heart is the cause of a *numerically* identical heart*,* i.e., our own, or a *qualitatively* identical heart, i.e., our offspring's heart.

Irrespective of whether or not function ascriptions entail (problematic) teleological explanations, there would be a crucial distinction between such explanations and behaviour explanations in terms of motivational responsiveness. *Firstly*, such behaviour explanations explain an *event* in terms of what that event is a factor of. In this respect, they differ from a teleological explanation of a *feature*, such as a heart, which explains it in terms of what that feature brings about.

Secondly, a behaviour explanation in terms of motivational responsiveness, in contrast to explanations of functions, explains *a token* event in terms of that token being a factor for *s.* That is to say, it does not (primarily) explain it in terms of historical tokens of this type *or* in terms of what this type of event tends to bring about. What this token event is a factor of depends, largely, on the concrete features of the dynamic environment in which it takes place. A teleological explanation of a functional event, by contrast, would merely explain a *token* event *in terms of what tokens of its type previously brought about or what its type tends to bring about.* A teleological explanation of a token functional event would explain *independently* of the fact that this token process is a factor for *s* to occur.

For these two reasons, and in particular the latter, we can distinguish '... in order to ...' explanations of action from other sorts of teleological explanations. Hence, we can safely assume that the above account of rational action in terms of motivational

responsiveness is mind-independent and generous – but not so generous as to fail to distinguish mere happenings and functional processes from rational actions.

3.5 Conclusion

In this chapter, I have developed the view that an action, x, gives application to a why-question for reasons *if* its progression depends on the fact that x is a factor for s to occur. The why-question for reasons for action is a request to explain the action on the ground of such a fact. I developed the thesis that such a fact is a cause of rational behaviour only if, in addition, an agent is motivationally responsive to it. I developed in addition the thesis that actions which give application to the why-question, thus conceived, might be independent of mental motivational responsiveness to what the action is a factor of. I furthermore argued that behaviour explanations differ from mere causal explanations and teleological explanations of functional events or reflexes. Therefore the 'factor of' account of action is mind-independent and generous, but not too generous. It allows us to characterize a class of mind-independent action which gives application to a why-question for reasons, yet allows us to distinguish such actions from mere events and functional processes.

4 – Motivational Responsiveness *or* Practical Judgment

4.0 Introduction

I am developing a *mind-independent*, generous account of rational action, because in order to develop not only the first part (i.e., that rational agency is independent of a capacity to consciously act for a reason) but also the second part (i.e., that the rational agency of an agent with agential self-consciousness is independent of a capacity to consciously act for a reason) of the dual capacity thesis, we need that account to be generous 'in the right kind of way'. The right kind of generous account would entail that the capacity to act for a reason is independent of mental reasons or consciousness of reasons. For unless it does entail this, it would be impossible to explain unconscious rational agency except as a feature of agents who lack agential self-consciousness. In the previous chapters I developed an account of motivational responsiveness to the fact that one's action is a factor for *s*, and emphasized that such responsiveness is mind-independent. In terms of mind-independence, this account seems to be heading in the right direction. Nevertheless, there is reason to doubt that this account is an account of *rational action*. As we will see, we philosophers ordinarily assume that an agent's doing *x* for reason *r* depends on his – implicit – practical judgment that *x* is *good* because of reason *r*. In addition, we assume that such judgments depend on consciousness: this

would strongly suggest that the 'factor of' account is either not mind-independent or not an account of rational action at all.

In this chapter I will argue against this ordinary assumption that doing *x* for reason *r* depends on a practical judgment that *x* is good because of reason *r*. In the first section I will introduce and develop an argument against what I call 'practical judgment theories' of rational action. I will develop the thesis that a capacity to act for reasons is independent of a capacity to form practical judgments, because reasons for action are external to the mind and independent of norms to act for certain reasons rather than others. In the second section I will develop the thesis that the question of whether a rational agent is capable of making practical judgments boils down to the question of whether he has specific mental capacities. I will infer from this that we can develop the thesis that the capacity to act for a reason is independent of a mental capacity, and we argue that the 'factor of' account of action is a mind-independent, generous account of rational action.

4.1 Acting for a Reason vs Acting in Light of a Norm

Traditionally, philosophy of action is wedded to what I call 'practical judgment theory'. On a *practical judgment theory* view, an agent is capable of doing *x* for reason *r iff* he has the capacity to make the practical judgment that *x* is 'good' on the ground of reason *r*. Such practical judgment theories are implicated by accounts of actions in terms of goals, values or norms (cf. Pink and Stone 2004; Millgram 2008). Such accounts of action *x* would equate *x*'s being conducive to a goal, or *x*'s aligning with values or norms as a reason for *x*.

Aristotle, for example, argued that action is aimed by an agent at an objective or subjective good (cf. Kraut 2012; Balaban 1990; for further reading see Pearson 2012). In more modern times, the good has been de-objectivized, and reasons for action were thought to derive from that action's coherence with subjective motivational states. Hobbes, for example, argued that we are basically rational egoists: we have desires, i.e., we value things subjectively and aim to satisfy our desires within the constraints of self-preservation (cf. Gert 1996; Homiak 2000). Others stressed that we are not really selfish egoists because, although not constrained by an objective good, we have social urges. Hume, for example, argued that we have a moral sense (cf. Mackie 2005, 44–74). Kant argued that we cause our actions on the basis of 'maxims', i.e., subjective principles of action which, he emphasized, we adopt in the normative light of formal

principles for making practical judgments (cf. McCarty 2009, 1–30). These principles give us reason for action. It has been emphasized that formal principles of rationality are not material enough to guide us (cf. Westphal 1991; Sedgwick 1988; Habermas 1988b). Hegel, for example, argued that the principles of rationality emerge in concrete socio-historic conditions which enable rational agency (Pippin 1991). This was followed by a more radical 'rage against reason' by existentialist and postmodern philosophers who, although they argued that our agency is independent of rationality *or* objective normative standards, seem committed to rationalizing an action in terms of choices, goals *or* values of agents (Crowell 2010; Solomon 1992). Other philosophers continued the more traditional line of action explanation. Some think of action along Aristotelian lines, in terms of 'natural goodness' (cf. Anscombe 1958; Foot 2001; Nussbaum 1988). Some continue to endorse that we are rational egoists (Gauthier 1987), others, that we have a moral sense (Kauppinnen Forthcoming). Others argue that the rules our actions depend on are normative rules which specify and shape our practices (Rawls 1955; Schapiro 2001; M. Thompson 2008). In a Kantian vein, Christine Korsgaard argues that we determine our action by adopting a principle for action in light of our practical identity, i.e., the basic description under which we value ourselves (Korsgaard 2009; Korsgaard 1996). Others argue that we determine our actions in light of norms to which we are committed as a result of intrinsically social practical deliberation (cf. Habermas 1988a; Apel 1975). Still others continue to emphasize that we act in light of our desires and higher-order desires or cares (Frankfurt 2004; Frankfurt 1971) or the strong evaluative commitments by which we mark our identity as socio-historical agents (Taylor 1985).

It might sound somewhat artificial to claim that these accounts all depend on one single explanatory scheme, i.e., a practical judgment theory of action – let alone on a practical judgment theory of rational action. Nevertheless, in my sense of the terms 'reason' and 'rational action', this is exactly what they do implicitly. Even those who 'rage against reason' would explain actions in ways which prove that such actions give application to a why-question for reasons. These sorts of reasons supposedly rationalize an action as something which happens because it is a *factor* for realizing a goal or an object of value or because it is a factor for satisfying a rule – quite irrespective of whether or not these goals, values or rules depend on rationality or are consciously adopted.

Aristotle assumed that action is regulated by a capacity to draw practical conclusions on the basis of reasons. In Aristotle's view, an agent who does *x* does so because he has

established (i) that *s* is good/desirable, (ii) that *x* is conducive to *s* and (iii) he drew the practical conclusion that he is to *x* (cf. Santas 1969). The modernist philosopher, because he de-objectivized the good, required a change in this Aristotelian concept. He assumed that an agent's doing *x* depends on a practical inference which an agent derives in light of his subjective motivational states (cf. Setiya 2004). Kant, in turn, emphasized that action depends not merely on subjective motivational states, but primarily on subjective principles for action which are subject to formal norms of rational action. Hegel diverted partly from this Kantian picture of agency. In his view, we act on the basis of rules which mark the socio-historical dimension of rational action (Pippin 1991). After Nietzsche, some philosophers emphasized that our capacity for action is basically pre-reflective and independent of objective goals or rational norms for action. Heidegger, for example, emphasized that 'engaged agency' is the basic dimension of our existence. On Heidegger's view, reflection is to be understood as a mode in which consciousness addresses obstacles which confront us as we are otherwise pre-reflectively engaging in our practical life-world (Taylor 1993). Existentialists argued that we have to make what Taylor calls a 'radical choice' so as to design our identity in the absence of pre-existing values, independently of any objective standards (Crowell 2010 in particular section 3).

Although, clearly, there has been and continues to be a debate about the kinds of goals, values and norms on which action depends, it is equally clear that action continues to be explained in terms of an agent's implicit capacity to resolve to do *x* because it is conducive to his goals or his values or other normative commitments. As a result, when asking the why-question with regard to action *x*, we expect an explanation of *x* which rationalizes *x* via the assumption that – implicitly – an agent exercised that capacity.

We must divert from this traditional understanding of the why-question if we are to develop a mind-independent, generous account of rational action. The motivational responsiveness account of action, which I developed in the previous chapter, allows us to do this. On this account, we may argue that the why-question for a reason is a request for a fact of the type: *x* is a factor for *s.* Such facts are possible independently of goals, values or norms. The fact that *x* is a factor for s does not speak for or against doing *x*. Yet, if an agent does *x* because *x* is a factor for s, then the fact that *x* is a factor for *s* is the *reason r* for which he does *x*. His capacity to *x* because of *r* might be *independent of* implicitly assigning a value to *x* on the basis of goals, values or norms. That is to say, doing *x* for reason *r* is to be understood independently of *assessing x*

relative to goals, values or norms. To appreciate this, witness the following imaginary scenario.

> A tired toddler asks his father, 'Why are you drinking coffee?' His father responds, 'I am drinking coffee in order to stay alert.' Then the toddler pours himself a coffee and starts drinking it. His mother enters the room and asks the toddler, 'Why are you drinking coffee?' The toddler responds, 'I am drinking coffee in order to stay alert.'

Obviously, there is a difference between the father's and the toddler's response. There is a difference not because they cite *different* kinds of facts (which they do not), but because they cite reasons *differently*. The father cites the reason to clarify what he is motivationally responsive to – independently of any goal, value or norm. By contrast, the toddler cites the reason in light of norms, in order to justify his drinking coffee.

Reasons *explain* actions only in light of an agent's motivational responsiveness to them. Reasons *favour* actions only in light of goals, values or norms. Independently of these, a reason for *x* counts neither for nor against doing *x*. A rational action, *x*, has the intriguing feature that it can be explained on the ground of an agent's motivational responsiveness to the fact that *x* is a factor for *s*, which (i.e., the fact that *x* is a factor for *s*), in light of goals, values or norms, speaks for or against his *x*-ing. It is important to notice that such a fact neither inherently explains nor justifies. To explain an action, *x*, in terms of such a fact, we must postulate an agent who is motivationally responsive to that fact. To justify *x* in terms of such a fact, we assume purposes, values or norms which justify doing *x*, because *x* is a factor for satisfying them.

In both cases we can call such facts 'reasons'. There is no need to hold that a reason for *x* counts in favour of *x*, nor that doing *x* because of a reason depends on an implicit practical judgment that *x* is good because of that reason.

Bittner is with me here (Bittner 2001, 131–44). As mentioned in the previous chapter, he argues that

> [...] to be a reason for which an action is done is to be something to which the action is a response. (Bittner 2001, 66)

In his view, acting for a reason is a matter of acting in ways which respond to the states or events. So, for example, an agent's being/getting tired is something to which an agent can respond. If, for example, he is drinking coffee in response to his being/getting tired, then he would be drinking coffee *for the reason that* he is tired/getting tired. On Bittner's view it is easy to see why reasons for action and acting for them are independent of normativity. Reasons are just states or events to which an agent may act in response; acting for reasons is just acting in response to these states or events.

In terms of developing a generous account of rational action, I am largely sympathetic to this account of reasons and rational action, for it entails that reason are external to the mind and that acting for them might be mind-independent. Yet I must disagree with Bittner on two scores. Firstly, the 'factor of' account of reasons and action differs from his account. This, of course, makes it difficult for me to rely on Bittner for developing the thesis that rational action is independent of practical judgment. Secondly, I am relying on a concept of action *x*, according to which a reason-explanation of *x* is, at heart, a '... in order to ...'-concept of *x*. Such a concept would imply that a reason-explanation of *x* explains *x* in terms of a specific sort of fact*: that x* is a factor for *s*. Bittner distances himself from any teleological explanation of action, suggesting that reason-explanations have no '... in order to ...' component. As a result, the 'factor of' account of rational action – in contrast to Bittner's – lends itself more easily to an interpretation in terms of the practical judgment theory. Given that it explains action in terms of '... in order to ...'-concepts, the 'factor of' account of rational action might be taken to imply that reason *r* for action *x* translates as *x* is a *means* to an *end* – *s*. This would suggest that an agent's doing *x* for a reason, *r*, depends on his doing *x* because *x* is a *means* to *s*. Hence, it suggests that this agent *aims for* or *values s* and does *x* because he assigns an instrumental value to *x* in light of *s*.

Thus, unlike Bittner, I have to prove that the 'factor of' account of action is independent of the practical judgment theory. To do so, I will basically draw on Bittner's argument that reasons for action are independent of norms. This argument supports his account of rational action as much as it would support mine, as it is mainly a *negative* argument *against* Korsgaard's and Broome's accounts according to which having ends has normative implications.

 Bittner summarizes Korsgaard's position as follows:

one's having a reason and one's being required to act accordingly both result from one's own legislative act. Reason and requirement are twins: born at the same time from a rational agent's setting herself an end. (Bittner 2001, 137)

In Korsgaard's view, reasons and norms are interdependent because they depend on an agent's setting an end for himself. According to her view, setting an end is a matter of adopting ends in light of a rational requirement to take the (necessary) means to achieve it. For example, an agent may set himself the goal of staying alert. Achieving this would depend on ensuring that he uses the necessary means to stay alert, which might, in a particular case, be drinking coffee. Bittner objects that Korsgaard's account of acting for reasons is too demanding. Why analyse an agent's having a reason to x in terms of his requiring himself to act for it?

On the 'factor of' view, having a reason to x is not only independent of imposing requirements on oneself, but also of having or adopting purposes. Having a reason for x is a matter of being in factual circumstances which determine what doing x is a factor of. That is, if the circumstances are such that doing x is a factor for s, then an agent who is in those circumstances has a reason to x irrespective of whether or not s is his goal or something he values or whether or not he is required to do anything.

Like Bittner, John Broome objects to Korsgaard's account of normativity that

Willing the end does not give you a reason to take the means, and it does not need to. So actually Korsgaard's conclusions do not follow. Willing an end need not give the end a normative status for you. Moreover, you can will an end without taking it as a law for yourself. You can simply decide to pursue it on one occasion. (Broome 1999, 418)

Nevertheless, Broome argues that having an intention to realize an end comes with a rational requirement to intend the means to it.

In general, intending an end normatively requires you to intend what you believe to be a necessary means. It does not give you a reason to intend what you believe to be a necessary means. (Broome 1999, 410)

Broome's account is subtle. In his view, rationality requires an agent to have certain propositional attitudes, depending on which other propositional attitudes he has. For

example, rationality requires an agent, who *intends* to stay alert and *believes* that drinking coffee is a necessary means to that effect, to *intend* to drink coffee. Broome emphasizes that such a requirement is independent of reasons and norms: such a requirement holds irrespective of whether an agent has a reason or is normatively required to (i) intend to stay alert, (ii) use the necessary means to do so or (iii) be rational (Broome 2007; Broome 2005; Broome 2002; Broome 1997). Broome would, for example, imply that it is irrational to *intend* to buy a boat, *believe* that borrowing money is a necessary means to buying a boat and still not intend to borrow money. Bittner objects to this, and argues that the agent who has adopted the aim of buying a boat and who believes he would be able to buy the boat only if he borrowed money might not borrow money but still not violate any requirement.

> If under circumstances as before you have the aim of buying a boat and do not borrow money, perhaps there is nothing at all wrong with you. Perhaps you are neither definitely failing in what you ought to do, nor failing provided that there is not something else you ought to do. Perhaps you are just fine. Sure, compared to people around you, you may be behaving strangely, but that is a different thing. Also, people may call you irrational. Yet if this only says that you are one of those who do not do what they have reason to do, this is true and admitted, but harmless, whereas if it implies that you are not as you ought to be, the comment is, for all we have heard so far, unwarranted. Broome's argument, then, presupposes, but it does not show, that reasons are normative. (Bittner 2001, 140)

Bittner's criticism of Broome is, in my opinion, slightly misleading. He criticizes Broome on the wrong ground. Broome would not argue that the agent violates a rationality requirement, but merely that he does not live up to standards for rationality. That is to say, on Broome's view the agent would violate a requirement only if he was required to act rationally. What is more, on Broome's view, rationality requirements for practical propositional attitudes are independent of practical norms or reasons for action.

> Because intention reasoning is reasoning, we may say it is normatively guided [...]. But intention reasoning is normative in no other way. Its content is not normative; it is not about what you ought to do or have a reason to do. (I use the term 'a reason' for a pro tanto reason. If you have a reason to do something, that means you ought to do it unless you also have a contrary reason not to.)

> [...] Furthermore, intention reasoning is not ought-giving nor even reason-giving [...] In my example, intention reasoning takes you from your intention of buying a boat and your belief that borrowing money is the only means of doing so, to an intention to borrow money. But it does not determine that you ought to borrow money, nor even that you have a reason to borrow money. (Broome 2002, 90)

We could take Bittner's objection to imply that the agent does not even violate a rationality requirement if he does not intend to borrow the money. However, in my opinion, endorsing this reading of Bittner's objection would not only be unconvincing, but would also distract us from the fact that doing x because x is a factor for s might not depend on intending to s at all. The central question for the present section was whether reasons are normative, not whether there are rationality requirements for practical propositional attitudes. The former question would be independent of the latter if reasons are not propositional attitudes and acting for reasons is independent of practical propositional attitudes. On the 'factor of' account of reasons, having a reason for some action is a purely factual matter, i.e., independent of the sort of propositional attitudes such as intentions and means-end beliefs which Broome argues are subject to rationality requirements. Hence, I resist the suggestion that reasons for action are dependent on the sort of rationality requirements which hold for having such attitudes. What is more, I deny that motivational responsiveness to reasons depends on such propositional attitudes. Hence, I conclude that Broome's account of rationality requirements does not prove that reasons for action and acting for reasons depend on norms for action.

Therefore the fact that x is a factor for s, as well as doing x because of that fact can both be understood independently of norms. Hence, there is no reason to explain action x for reason r via the assumption that an agent who does x – implicitly – judged that, because of r, he is to do x.

4.2 Conceiving Acting for Reasons

To understand more fully the view that rational action is independent of practical judgment, let us imagine a room in which an agent, A_1, is drinking coffee in order to stay alert, a cat, A_2, is walking to its bowl in order to get some food, and another agent, A_3, conceives what is going on. This scenario involves four ways of conceiving acting for reasons.

(I) A_1 observes that A_2 is walking to its bowl in order to get some food;

(II) A_3 observes that A_2 is walking to its bowl in order to get some food;

(III) A_3 observes that A_1 is drinking coffee to stay alert;

(IV) A_1 non-observationally conceives that he is drinking coffee to stay alert.

These cases can be classed in the following way:

(A)   Spectator S_n conceiving that an agent, A_n, does x in order to s (cases I, II and III);

(B)   An agent's (A_n) non-observationally conceiving that he does x in order to s (case IV).

There is a difference between (I) and (II), on the one hand, and (III), on the other hand. The spectator conceives that the cat is walking to its bowl in order to get some food (i.e., because walking to his bowl is a factor for getting some food), but would not explain the cat's action as involving – implicitly – practical judgments. He will, normally, not assume that the cat walks to his bowl because he assigns value to his walking to his bowl as a means to get food. The spectator may hold that the cat's behaviour does not live up to certain standards for action, say the one of prudence. For example, when the cat has a tendency to eat everything in his bowl in one go, as a result of which he remains starving in the evening, the cat can be said to act against its own interests. The spectator will, however, not assume that the cat is normatively required to be prudent and stop eating everything in his bowl in one go.

By contrast, if the spectator conceives the coffee-drinker's action in terms of this standard for prudence, then he will assume that the coffee-drinker can and should act on that standard. For example, if the coffee-drinker has a tendency to drink too much coffee during the day so that he has sleepless nights and is tired in the mornings, then the spectator will assume that the coffee-drinker acts against his own interest and that prudence tells him to stop drinking coffee.

Thus, a spectator's conception of rational actions might have three aspects. He conceives action x for reason r

(i)   in terms of an agent's motivational responsiveness to reason r;

(ii)   in terms of standards for action;

(iii)   on the assumption that the agent should act on these standards.

The spectator might conceive the cat's and the coffee-drinker's actions solely as (i) or also as specified in (ii). However, if he conceives their behaviours as (ii) specifies, then he is quite likely to emphasize that the coffee-drinker's behaviour should, but the cat's should not, be conceived as specified in (iii). Why is this so? Why do we assess the coffee-drinker's, but not the cat's behaviour, as something which should abide by such norms?

To explain this difference, we must track a capacity which the spectator assigns to the coffee-drinker, but not to the cat: a capacity by which he is not only capable of, but is also responsible for acting on the basis of these norms. However, why does the spectator assign this capacity to the coffee-drinker and not to the cat? Is there a difference between the cat and the coffee-drinker, based on which the spectator would be justified in this? Of course, we can assume that the coffee-drinker, but not the cat, has the capacity or the responsibility to act on these standards because he – but not the cat – is subject to a norm to act prudentially. Assuming this would, however, yield a largely circular answer to the question at stake. The question is why we conceptualize the coffee-drinker's, but not the cat's, behaviour in terms of a requirement and a capacity to act prudentially. To answer this question we must explain why – and not assume that – we impose the requirement on the coffee-drinker and not on the cat.

Both the cat and the coffee-drinker are capable of acting for a reason. So we cannot explain the difference between the cat's and the coffee-drinker's capacity and responsibility in terms of that capacity. We must therefore explain it in terms of other capacities which the coffee-drinker has, but the cat lacks. One distinctive feature of the coffee-drinker is that he non-observationally conceives that he is drinking coffee to stay alert. This form of conceiving was referred to in (IV). Consequently, to rationalize the spectator's holding the coffee-drinker responsible, we might argue that the coffee-drinker's capacity to conceive that he is acting for a reason depends on his capacity and responsibility to act in light of standards. Alternatively, we may argue that he is capable of acting for norms, because of other features.

Whichever approach we take, we could be assuming that a capacity to act for a reason is independent of a capacity to make practical judgments.

Hence, we can safely conclude that on the 'factor of' account, a capacity to act for a reason might involve, but is independent of, a capacity to make practical judgments.

This, in turn, establishes that the 'factor of' account of a capacity to act for a reason is generous in the right kind of way: it accounts for the possibility that a capacity to act for a reason is not a mental capacity. Therefore, we can conclude that we have fulfilled one of the requirements for arguing that mere agential self-consciousness is possible. That is, we have developed the first part of the dual capacity thesis: a capacity to act for a reason is independent of a capacity to consciously act for a reason. Therefore, in order to develop the extraordinary thesis about agential self-consciousness, we need only to develop the second part of the dual capacity thesis: a capacity to make explicit that one is acting for a reason is independent of a capacity to consciously act for a reason. Developing this second part will be the task in the subsequent chapters 5 and 6.

4.3 Conclusion

In this chapter I have argued that the 'factor of' account of action in terms of motivational responsiveness to reasons is a mind-independent, generous account of rational action. As a result, I proved that the extraordinary thesis about agential self-consciousness could be developed if we developed the thesis that a capacity to make explicit that one is acting for a reason is independent of a capacity to consciously act for a reason. In the *first* section I developed the thesis that a reason for action does not inherently count in favour of that action and that acting for a reason is independent of favouring an action on the ground of a reason. In the *second* section, I introduced various ways of conceiving an action in terms of reasons. From a spectator's standpoint, we basically conceive rational actions in two different ways, depending on whether or not we assume that an agent is responsible for and capable of acting on certain practical requirements. From this I inferred that the capacity to act for a reason is independent of a capacity to make practical judgments. I concluded that a capacity to act for a reason would involve a capacity to consciously act for a reason only if combined with an additional capacity which depended on a capacity to consciously act for a reason. This proved that the account of action in terms of motivational responsiveness can be called a mind-independent, generous account of rational action.

# 5 – Making Explicit that One is Acting for a Reason

5.0 Introduction

Is an agent's capacity to make explicit that he is acting for a reason independent of a capacity to consciously act for a reason? On the face of it, *making explicit that one is drinking coffee to stay alert* involves agential self-consciousness much more directly than does drinking coffee to stay alert. In the previous chapters I have developed a mind-independent account of rational acts such as drinking coffee to stay alert. The key feature of this account is that rational agency might merely involve mind-independent motivational responsiveness to what actions are factors of. However, I have not addressed the question of whether the sort of activity which enables agential self-consciousness depends on a capacity to consciously act for a reason. In other words, the mind-independent account does not prove that making explicit that one is acting for a reason is independent of a capacity to consciously act for a reason. Hence, I must develop the thesis that it is, if I am to develop the thesis that agential self-consciousness is independent of a capacity to consciously act for a reason. In the present chapter, I will develop the thesis that the capacity to make explicit that one is acting for a reason is

independent of a capacity to consciously act for a reason. In the subsequent chapter I will develop a model of agential self-consciousness in terms of that thesis.

The capacity to make explicit that one acts for a reason is part and parcel of a tripartite capacity to act for a reason. This tripartite capacity comprises a capacity

(i)   To φ for reason r, i.e., to perform an action for a reason;

(ii)   To ψ for reason s, i.e., to ask a why-question *in order to* request a reason-explanation of an action that one is performing; and

(iii)   To ρ for reason t, i.e., to make explicit that one is acting for a reason *in order to* give a reason-explanation of an action that one is performing.

Like any other action, ψ-ing *and* ρ-ing can be rational. This means that we might conceive both of them as actions about which we can raise why-questions and about which we can make explicit that an agent does them for a reason. In other words, ρ-ings and ψ-ings can be conceived as φ-ings, on which further ρ-ings and ψ-ings depend. And, of course, these latter ρ-ings and ψ-ings can be considered as φ-ings, on which still further ρ-ings and ψ-ings depend, etc. We get the following schema.
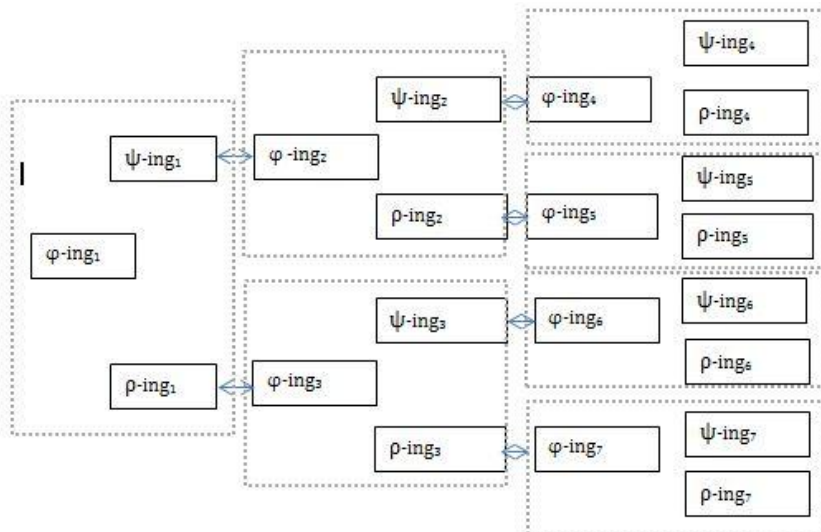
In this schema I assigned indexes to φ, ψ and ρ. In the basic case the index of φ is n=1. The index n of that φ is inherited by ψ, for which it asks for a reason-explanation.

Action ρ inherits the index of that φ, for which it gives a reason-explanation. Considered as φ, ψ has double the index which it has as ψ. Considered as φ, ρ has double the index it has as ρ plus 1.

I will develop the thesis that the capacities to φ, ψ and ρ for a reason are all independent of a capacity to consciously φ for reason r. In the first section, I will develop the view that a capacity to ρ_n is *independent of a capacity to consciously ρ_n for a reason.* In the second section, I will develop the thesis that the tripartite capacity on which the capacity to ρ depends is independent of a capacity to consciously φ for reason r.

5.1 Exercise of a Capacity to Consciously Act?

Kant would argue that an agent's capacity to ρ, i.e., to make explicit that he is acting for a reason, depends, as a practical capacity, in two ways on a capacity to consciously act for a reason. On Kant's view, an agent's capacity to act for a reason is his will, i.e., his practical reason (Kant 1999 chapter two). Hence, an agent who says, 'I am φ-ing for reason r' entails not only 'I *will* to φ for reason r' but also that he wills to make explicit that he φ's for reason r. Hence, he implicitly says, 'I *will* to ρ for reason t' (cf. McCarty 2009 in particular the first chapter). In addition, and more generally, the 'I' by which an agent identifies himself and to which he attributes his φ-ing and ρ-ing is, on Kant's view, self-reflexive conscious activity and, consequently, *rational* self-conscious activity (Pippin 1987, 473–5).

To understand these two aspects of Kantian philosophy, we might regard them as his critical response to two approaches in philosophy: empiricism and rationalism (McCormick 2005; cf. Markie 2012; Norton 1981, p.331). Empiricist approaches have it that our *senses* provide for ideas and passions*,* and that *reason* analyses and compares these ideas and passions. Rationalist schools argue that reason provides for knowledge independent of the senses. Kant argues that the *senses* and *reason* are interdependent and are best understood as aspects of spontaneous, self-reflexive, judgmental activity (cf. Williams 2009). On Kant's view, such judgmental activity is self-regulating: it provides for the principles on which it judges – i.e., principles for representational, practical or aesthetic judgment. The way I would put it is that a self-regulating, judgmental activity such as this involves a capacity to consciously act for a reason. That is to say, it involves a conscious capacity to judge that such and such is so because judging in this way is a factor for satisfying a principle of judgment.

Kant shares with the empiricists and rationalists a conceptual schema of the mind which is incompatible with the extraordinary thesis about agential self-consciousness. That is to say, he agrees with these schools that mental states are products of mental faculties, sense and thought. However, he endorses the idea that sensory representations depend on a reflexive, and therewith rationally constrained, judgmental subsumption of sense impressions under categories of thought.

Consequently, on the one hand, I need to distance myself from Kant's positive accounts of agency as well as his account of consciousness in general. On the other hand, I cannot withdraw to empiricist or rationalist models of agency and consciousness, as both models seem to explain agency and consciousness in terms of reason, i.e., that there is *minimally* a conscious capacity to analyse, or a capacity for means-end reasoning. Such capacities would be versions of a capacity to consciously act for a reason, as they would be conscious capacities to draw theoretical or practical conclusions from reasons.

I have already formulated an account of action in the previous chapter which would allow us to claim that rational agency is independent of a capacity to consciously act for a reason and, a fortiori, independent of presumed mental faculties such as reason. In the present and subsequent chapters I will develop a concept of consciousness that is independent of a capacity to consciously act for a reason. In general, I will divert from empiricist, rationalist and Kantian models of consciousness, assuming that 'I' refers to a *non-mental* entity with a mind and self-consciousness. I will argue that we can explain certain features of and processes in an agent's mind, including his self-consciousness, in terms of his (rudimentary) non-mental capacity to use language.

In the remainder of the present and in the subsequent chapter, I will develop the thesis that 'I' does not refer to consciousness, but to an entity which, like the Strawsonian person, is the logical subject of mental as well as non-mental predicates (Strawson 1959, 87–116). In light of this view, I will develop the thesis that an agent's capacity which explains conscious states and events such as self-consciousness, thought and thought processes can be conceived as a person's non-mental capacity to use a *public* language.

I take Wittgenstein's private language argument as an occasion to launch my thesis (Wittgenstein 2003 §§ 244-71; Candlish and Wrisley 2012). In his critical assessment of traditional philosophy of meaning and language, Wittgenstein argued that the idea of a

private language is incoherent. That is to say, it is impossible for an individual to classify his own experiences if his experiences are merely private. Hence it follows that self-consciousness, 'the self' and its mind, are not private. This Wittgensteinian argument is essentially negative, i.e., a rejection of a particular concept of the nature of the objects which can be conceived. Nevertheless, it suggests that self-consciousness depends on an agent's mental states being public and on his capacity to classify his mental states in terms of a 'public language'. Hence, he and his mental states must have public features, which alone allow us to refer to ourselves and attribute mental states to ourselves.

Wittgenstein's approach to self-consciousness shares one important feature with the one which I am to develop. It suggests that socio-linguistic capacities which enable an agent to attribute a mental state to himself are somehow independent of an agent's consciousness that he has these mental states. It is important, of course, to understand the sense in which Wittgenstein suggested that the former is independent of the latter. One natural reading of this independence would be to take Wittgenstein to have asserted that the capacity to attribute mental states to oneself is independent of a merely *private* consciousness of these states. This reading, I think, fits behaviourists' and pragmatists' revisionary approaches to consciousness, according to which there is no place for a private mind. Hard-core behaviourism could suggest that this proves that there is no mind, but only public behavioural dispositions which can be attributed to an agent and which an agent can attribute to himself (Luckhardt 1983). Others, for example pragmatists, could take Wittgenstein to imply that the mind is public and that the capacity for attributing mental states to oneself depends on the fact that the mind is public and that these capacities are part a public mind to which it attributes these mental states (cf. Colapietro 2006; Haack 1982; Markell 2007).

I take Wittgenstein's argument as an opportunity to entertain a distinct and more radical thesis: the capacity to use a communal language is not a mental one, i.e., is independent of consciousness – be it private or public. What is more, this capacity allows an agent to attribute mental states to himself independently of a capacity to consciously act for a reason. If we could develop these theses, then we would have one vital ingredient for an argument that a capacity to make explicit that one is acting for a reason is independent of a capacity to consciously act for a reason. In the remainder of this section, I will critically discuss Mead's account of significant gestures to develop the thesis that a capacity to make explicit that one is acting for a reason is not itself a

capacity to consciously act for a reason (Mead 1925; Mead and Morris 1967 in particular section II and III; cf. Silva 2007). Then in the next section I will develop the thesis that this capacity is even independent of a capacity to consciously act for a reason.

A basic ingredient of Mead's account of significant gestures is his concept of the social act, which is as follows:

> A social act may be defined as one in which the occasion or stimulus which sets free an impulse is found in the character or conduct of a living form that belongs to the proper environment of the living form whose impulse it is. I wish, however, to restrict the social act to the class of acts which involve the co-operation of more than one individual, and whose object as defined by the act, in the sense of Bergson, is a social object. I mean by a social object one that answers to all the parts of the complex act, though these parts are found in the conduct of different individuals. The objective of the act is then found in the life-process of the group, not in those of the separate individuals alone. (Mead 1925, 263–4)

Mead's social act is social in a double sense. It has a social objective, and it belongs to cooperating individuals. Many forms of life are capable of such action.

Some socials acts not only have a social objective and belong to cooperating individuals, but are carried out on the basis of 'representations' of the parts that compose them. 'Such an act would be one in which the different parts of the act which belong to different individuals should appear in the act of each individual' (Mead 1925, 264). This latter kind of social act stands in contrast to those in which the cooperation is a *mere* 'exchange of gestures' between individuals, independent of a 'representation' of the complete social act.

All social actions are composed of *gestures*, i.e., 'that part of the act or attitude of one individual engaged in a social act which serves as the stimulus to another individual to carry out his part of the whole act' (Mead 1925, 270). Certain kinds of gestures, among which the vocal gesture, might arouse 'in the individual who makes it a tendency to the same response that it arouses in another, and this beginning of an act of the other in himself enters into his experience, he will find himself tending to act toward himself as the other acts toward him' (Mead 1925, 271). Such gestures are key to the possibility of

significant social gestures, i.e., gestures performed to arouse a response in order to complete a social act.

> Where a vocal gesture uttered by one individual leads to a certain response in another, we may call it a symbol of that act; where it arouses in the man who makes it the tendency to the same response, we may call it a significant symbol. These organized attitudes which we arouse in ourselves when we talk to others are, then, the ideas which we say are in our minds, and in so far as they arouse the same attitudes in others, they are in their minds, in so far as they are self-conscious in the sense in which I have used that term. But it is not necessary that we should talk to another to have these ideas. We can talk to ourselves, and this we do in the inner forum of what we call thought. (Mead 1925, 272)

There are vital differences between merely meaningful gestures and significant ones; I will mention two of the latter type. Firstly, A's significant gesture depends on A's anticipating the social response to *his* making that gesture. On Mead's view, an agent who is capable of anticipating the social response to his making a gesture is a *self*, '[…] that is, an individual who organizes his own response by the tendencies on the part of others to respond to his act (Mead 1925, 267). The self, thus conceived, is the object of self-consciousness. 'We appear as selves in our conduct in so far as we ourselves take the attitude that others take toward us, in these correlative activities' (Mead 1925, 268).

Secondly, responding to a significant gesture is a matter of *agreeing* or *disagreeing* with it. That is to say, an agent who makes a significant gesture makes it in order to arouse in others the same response that it arouses in him. Consequently, if others respond in the same way that he does, then they agree with him. Otherwise, they disagree. As a result, an agent who makes a significant gesture (implicitly) makes it in order to arouse in others the agreement to the gesture that it arouses in him.

For the purposes of developing the extraordinary thesis about agential self-consciousness, I am fairly sympathetic to Mead's suggestion that gestures are devices for completing social actions. Furthermore, I am supportive of his thesis that ideas, thoughts and self-consciousness *depend* on a capacity to use significant gestures. Nevertheless, in contrast to Mead, I must emphasize that the exchange of significant

gestures occurs logically and explanatorily prior to the ideas, thoughts and self-consciousness which it might enable. More explicitly, the capacity to exchange significant gestures might involve, but not depend on, self-consciousness at all.

To clarify my revisionary reading of Mead, let us briefly return to Mead's distinction between *meaningful* and *significant* gestures. Social action is, according to Mead, inherently *meaningful* in the sense that a social action is an exchange of gestures between agents A and B: agent A makes an initial gesture and agent B makes a gesture in response to it, and this exchange continues until a social objective is achieved. Agent A's (respectively B's) gesture is *meaningful* as it arouses in B (respectively A) an attitude to perform his part in a social action. On my reading, this exchange of gestures may be wholly unconscious. A's capacity to make an initial gesture may be independent of a conscious capacity to make that gesture, of consciousness of the attitudes which that gesture arouses in B, and of consciousness of any social objective. B's response to A's initial gesture, in turn, may be independent of any consciousness of A's gesture, of a conscious attitude to respond, of consciousness of attitudes which his response arouses in A, and of consciousness of the social objective.

A capacity to make significant gestures – in contrast to the capacity for consciously making meaningful or significant gestures – can be understood as a practical capacity to (i) *covertly* make a gesture and arouse in oneself the social response to it before *overtly* making a gesture *and* (ii) overtly make a type-like gesture that depends on the social response aroused by the gesture covertly made. Such a capacity might as well not be a capacity to consciously act for a reason.

Far from rejecting Mead's pragmatist model of significant gestures, I adopt it for present purposes *but* emphasize that the capacity to make significant gestures need not be a conscious one – irrespective of whether or not it enables self-consciousness. Therefore, on my reading of Mead's account of significant gestures, we might well analyse an agent's capacity to make explicit that he is acting for a reason *as his* capacity to make a gesture to arouse in others what it aroused in him: agreeing that he is acting for that reason. If we analyse an agent's capacity to make explicit that he is acting for a reason in this way, we would be able to argue that the capacity to ρ_n is independent of a capacity to consciously ρ_n for a reason.

Does this capacity to ρ *depend* on a capacity to consciously φ for a reason?

5.2 Dependence on a Capacity to Consciously Act?

Does an agent's capacity to make explicit that he is acting for a reason depend on a capacity to consciously act for a reason? In the introduction to this chapter, I claimed that an agent's capacity to make explicit that he is φ-ing for reason r *depends* on a tripartite capacity to act for a reason. He is capable of

(1) φ-ing, i.e., performing a rational action;

(2) 'ψ-ing', i.e., asking why he is φ-ing; and

(3) 'ρ-ing', i.e., making explicit that he is φ-ing for reason r.

Hence, the capacity to make explicit that one is acting for a reason involves a tripartite capacity to raise and answer why-questions about φ-ings like drinking coffee, but also about φ-ings like raising and answering why-questions. The question to be settled is whether such a tripartite capacity depends on a capacity to consciously act for a reason. That is, whether it entails that (i) for n=1, a capacity to φ_n, ψ_n or ρ_n is a capacity to consciously act for a reason or that (ii) φ_{n+x}, ψ_{n+x} or ρ_{n+x} involves a capacity to consciously act for a reason.

Witness the following imaginary – and somewhat artificial – conversations between our coffee-drinking agent, A, and some other agent, B.

Case 1

Agent A says: 'I am drinking coffee to stay alert.'

B responds: 'Why are you saying that?'

Continuation a) A responds: 'Why are you asking, "Why are you saying that?"'

B responds: 'I am asking you to give a reason- explanation of your saying it.'

Continuation b) A responds: 'In order to answer the why-question about my drinking it.'

Case 2

Agent A says: 'I am saying that I am drinking coffee to stay alert in order to answer the why-question about my coffee-drinking.'

B responds: 'Why are saying that?'

Continuation a) A responds: 'Why are you asking "Why are you saying that?"'

B responds: 'I am asking you to give a reason- explanation of your saying it.'

Continuation b) A responds: 'In order to answer the why-question about my saying it.'

Let me start with *three* observations about these two cases.

1) These two cases and their continuations are similar in structure, except that the second case starts with what appears to be a 'reflexive' exercise of a capacity to make explicit that one is acting for a reason. Agent A exercises that capacity to give a reason-explanation of his (earlier) exercise of that capacity. The first case, by contrast, starts with a 'linear' exercise of the capacity to make explicit that one is acting for a reason.

2) In both cases, making explicit that one is acting for a reason depends on a capacity to ask the why-question. Asking a why-question, in turn, addresses and thus depends on a capacity to make explicit that one is acting for a reason. Consequently, the capacity to make explicit that one is acting for a reason and the capacity to ask the why-question are – somehow – interdependent.

3) The capacity to make explicit that one is acting for a reason and the capacity to raise a why-question are interdependent; therefore the reflexive exercise of the former depends on a reflexive exercise of the latter.

Do the *reflexive* exercise of a capacity to make explicit that one is acting for a reason or the *interdependence* of a capacity to make explicit that one is acting for a reason and a capacity to raise a why-question, require an explanation in terms of a capacity to consciously act for a reason? In the remainder of this chapter, I will develop the thesis that they do not.

An agent's capacity to make explicit that he is drinking coffee to stay alert depends on his capacity to drink coffee to stay alert and his capacity to ask why he is drinking coffee. Does one of these capacities depend on a capacity to consciously act for a reason? The capacity to drink coffee does not, at least not in light of the generous account of rational action developed in previous chapters. This is so in a case involving an agent who has no capacity to make explicit that he is acting for a reason. Is it also the case if an agent has a capacity to make explicit that he is acting for a reason? I refer to a YouTube movie featuring a monkey smoking a cigarette (Anon. 2007). Ignoring for a moment issues relating to how it came to smoke – which might lead us to explain his behaviour via a conscious, practical capacity of human beings – let us imagine which additional capacities the monkey would need in order to be capable of making explicit that he is smoking for a reason. Would he need a capacity to consciously smoke for a reason? Would he need a capacity to make explicit that he is acting for a reason, which depends on a capacity to consciously act for a reason? Or would he merely need a capacity to make explicit that he is acting for a reason, which is independent of a capacity to consciously act for a reason (cf. Anon. 2008b)? On my view, it would be sufficient if he had a capacity to make explicit that he is acting for a reason – irrespective of whether or not this latter capacity depends on a capacity to consciously act for a reason.

If the capacity to drink coffee does not depend on a capacity to consciously act for a reason, then does the capacity to answer the why question depend on a capacity to consciously act for a reason? The capacity to raise a why-question and the capacity to make explicit that one is acting for a reason are interdependent in the sense that a specific exercise of one of them requires the presence of both capacities. An agent can ask himself, 'Why am I drinking coffee?' only if he has the general capacity to give a reason-explanation for what he does. Similarly, an agent will say, 'I am drinking coffee to stay alert' only if he has the general capacity to ask why he is doing what he is doing. Importantly, however, the capacity to raise the *concrete* question, 'Why am I drinking coffee?' as well as the exercise of that capacity are independent of a capacity to make
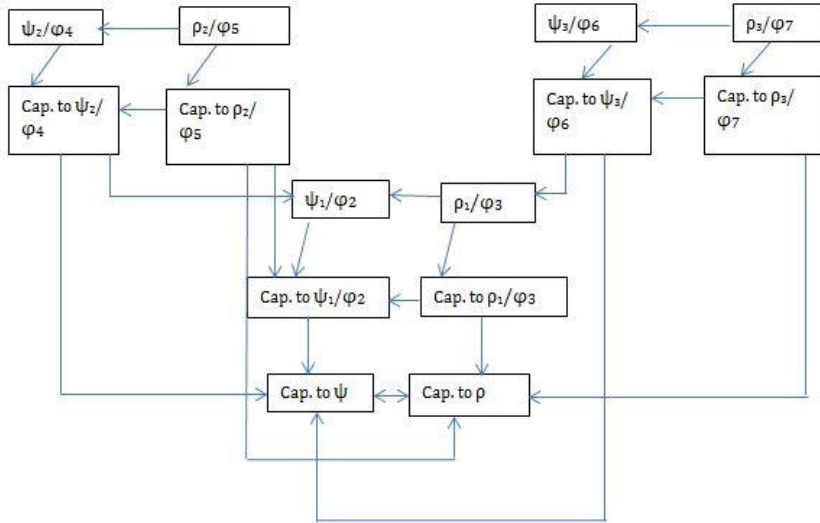
explicit that one is drinking coffee to stay alert. That is to say, an agent may well ask himself, 'Why am I drinking coffee?' and lack a capacity to make explicit that he is drinking coffee for a reason. By contrast, the capacity to make explicit that one is drinking coffee to stay alert depends on the capacity to raise the concrete question, 'Why am I drinking coffee?' Equally, the exercise of the former depends on the exercise of the latter. An agent will only say, 'I am drinking coffee to stay alert' if – implicitly – he said to himself, 'Why am I drinking coffee?' We get the following schema of dependencies. The arrow denotes a 'x depends on y'-relation, where it departs from x and goes in the direction of y.
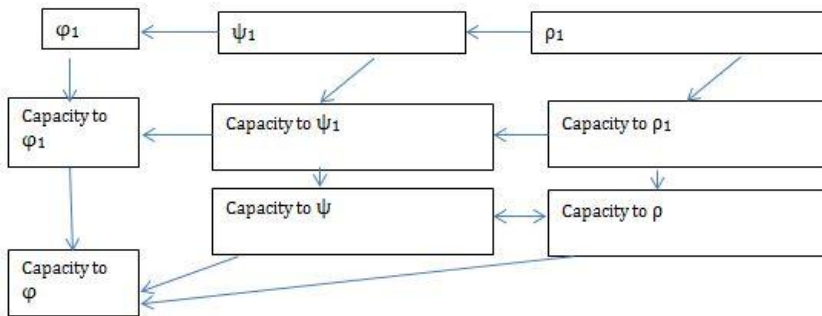
Does the capacity to make explicit that one is acting for a reason depend on a capacity to consciously act for a reason? The above schema would suggest that it does not, except if the capacities to ask 'Why am I drinking coffee?' or to respond 'I am drinking coffee to stay alert' did. Are they capacities to consciously act for a reason?

If the capacity to drink coffee to stay alert is independent of a capacity to do so consciously, then what reason is there to assume that asking for or giving a reason-explanation of one's drinking coffee depends on such a capacity? I can only imagine one reason: an agent's ψ-ing$_n$/ρ-ing$_n$ for a reason depends on his *general* capacities to ψ and ρ for a reason. By virtue of this, an agent who is ρ-ing$_n$ will quite likely be able to make explicit that he is ψ-ing$_n$ for a reason *as well as* that he is ρ-ing$_n$ for a reason. In this sense, then, the capacity to ρ for a reason *entails* a double capacity to φ_n for a reason, which depends on a capacity to ρ_n for a reason. On the other hand, there is a difference between *causing/entailing* something and *depending on* that thing. That is to say, the general capacity to ψ and ρ, and the specific capacities to ψ_n for a reason and ρ_n for a reason are *logically* and *explanatorily prior* to a capacity to make explicit that one is ψ-ing$_n$/ρ-ing$_n$ for a reason. We get the following schema.

The arrow denotes an '*x* depends on *y*'- relation, where it departs from *x* and goes in the direction of y.

ψ_2/φ_4 ρ_2/φ_5 ψ_3/φ_6 ρ_3/φ_7

Cap. to ψ_2/φ_4 Cap. to ρ_2/φ_5 Cap. to ψ_3/φ_6 Cap. to ρ_3/φ_7

ψ_1/φ_2 ρ_1/φ_3

Cap. to ψ_1/φ_2 Cap. to ρ_1/φ_3

Cap. to ψ Cap. to ρ

In this schema we find no sign of a capacity to consciously act for a reason, i.e., a capacity to make explicit that one is φ_n-ing for a reason which is not only entailed by, but is logically and explanatory prior to a capacity to φ_n for that reason.

Things might appear more complicated given the possibility that the capacity to ψ_2 and to ρ_2, on the one hand, and to ψ_3 and to ρ_3, on the other hand, might depend on still further capacities to act for a reason which, in turn, might involve a capacity to consciously act for a reason. However, we can easily exclude this possibility on the same structural grounds on which we excluded that the capacity to ψ_1 and ρ_2 depended on a capacity to consciously act for a reason. The specific capacities to ψ_{n+1}/ρ_{n+1}, enabled by a combination of general capacities to ψ and ρ, on the one hand, and specific capacities to ψ_n/ρ_n are not involved in these general and specific capacities. Hence, the former are independent of a capacity to consciously act for a reason. This answer holds, irrespective of whether n = 1, 2, 3 or 1000.

5.3 Conclusion

In this chapter I developed the claim that a capacity to make explicit that one is acting for a reason is independent of a capacity to consciously act for a reason. I have developed firstly the thesis that the capacity to make explicit that one is acting for a

reason may be a non-conscious capacity to state facts about oneself in terms of a public language. Secondly, I have developed the idea that a capacity to make explicit that one is acting for a reason is independent of a capacity to consciously act for a reason.

I conclude that, on this conception of it, the capacity to make explicit that one is acting for a reason is independent of a capacity to consciously act for a reason. In the subsequent chapter I will clarify how one can explain agential self-consciousness in terms of this capacity to make explicit that one is acting for a reason which is independent of a capacity to consciously act for a reason.

6 – *Mere* Agential Self-consciousness

6.0 Introduction

The aim of my dissertation is to develop the extraordinary thesis about agential self-consciousness: that agential self-consciousness is independent of a capacity to consciously act for a reason. To that effect, I have developed the thesis that a capacity to act for a reason *and* a capacity to make explicit that one is acting for a reason are *both* independent of a capacity to consciously act for a reason. To conclude my attempt to develop the extraordinary thesis, I will clarify in the present chapter that agential self-consciousness can be conceived as an effect on consciousness of the capacity to make explicit that one is acting for a reason.

This account of agential self-consciousness stands in some contrast to accounts of non-observational knowledge and to self-consciousness in general. In the philosophical literature we find two types of explanations of an agent's *non-observational* knowledge of his actions. In terms of Humberstone's distinction, we could either conceive such

consciousness as 'thetic' or as 'telic' (Humberstone 1992).[7] Both explanations offer little prospect for the extraordinary thesis about agential self-consciousness. The *telic* consciousness view would explain agential self-consciousness as an intrinsic feature of an agent's action (cf. Setiya 2008). The *thetic* consciousness view would explain agential self-consciousness as an agent's belief about his actions which, once combined with certain motives, would be causally involved in regulating action (cf. Velleman 2000; Paul 2009).

Self-consciousness has been explained on the assumption of special faculties of a mind by which the mind conceives itself. Such features range from a mind's inner sense to the transparency of mental states or reflexivity of mind (cf. Shoemaker 1994; Frank 2002; Moran 2004a). Such models of self-consciousness offer dim prospects for the extraordinary thesis about agential self-consciousness, at least insofar as they suggest that self-consciousness is primarily a mind's consciousness of its states, for this would suggest that agential self-consciousness is primarily a mind's consciousness of its practical states.

In this chapter I will sidestep such explanations of knowledge of action and agential self-consciousness. In contrast to the thetic view, I must develop the view that there is no need to assume that non-observational knowledge of actions is the sort of thing that causes action. In contrast to the telic view, I will have to develop an argument that there is no need to assume that an agent's consciousness that he has practical capacities depends on these practical capacities being mental/conscious capacities. In contrast to prevalent traditional models of self-consciousness, I will have to argue not only that – vitally – self-consciousness is consciousness of oneself as a non-mental entity with mental capacities, but also that the object of self-consciousness is publicly identifiable. Moreover, I will develop the view that a non-mental capacity to refer to oneself as such

---

[7] Here I discuss knowledge, and in particular distinctions pertaining to it, as a form of consciousness. The interrelations between consciousness and knowledge are not easy to spell out. There have been discussions about how some forms of knowledge are 'tacit' or 'implicit'(cf. Polanyi 1962) and about how there is a distinction to be recognized between 'know that' and 'know how' (cf. Ryle 1945; Stanley 2011; Bengson and Moffett 2012). These latter forms or knowledge have been tightly associated with processes or capacities which exist beyond our consciousness of them. However, irrespective of whether unconscious knowledge is possible, the extraordinary thesis about agential self-consciousness is a claim about self-consciousness. In this context I explore whether, mutatis mutandis, we might explain mere agential self-consciousness on the basis of concepts of *and* distinctions between forms of knowledge. In other words, assuming that one of these forms of knowledge enables agential self-consciousness, could such agential self-consciousness be independent of a capacity to consciously act for a reason?

an object enables self-consciousness. From this I will develop the view that agential self-consciousness is independent of a capacity to consciously act for a reason.

The remainder of this chapter is in *four* sections. In the *first* section I will discuss contemporary accounts of an agent's non-observational knowledge of his action, which explain such knowledge on the assumption that the causes of action are mental. Then in the *second* section I will briefly discuss approaches to self-consciousness which explain self-consciousness in terms of a subjective capacity to identify oneself. In the *third* section I will reconstruct Tugendhat's Wittgensteinian theory of self-consciousness, according to which a subject's knowledge that he has mental states depends on his capacity to indicate himself and attribute mental states to himself using a communal language. In the *fourth* and final section I will adduce Tugendhat's account to explain an agent's capacity to justifiably and independently of observation claim that he has *non-mental* practical states. Then I will explain agential self-consciousness as a publicly identifiable effect of this latter capacity on his mind. I will conclude that an agent can have agential self-consciousness independently of a capacity to consciously act for a reason.

6.1 Non-observational Knowledge

In the first chapter of this dissertation, I introduced Anscombe's account of intentional action, according to which an agent's intentional action depends on his practical knowledge of it. Such practical knowledge of his action would, on Anscombe's view, which I discussed in chapter 2, be the way in which an agent non-observationally knows what he is doing (cf. Falvey 2000; Moran 2004b).

Some philosophers agree with Anscombe that an agent must *non-observationally* know what he is doing intentionally. Those philosophers disagree, however, on how to explain such knowledge. One basic disagreement concerns the question of whether or not such knowledge is practical or whether it is thetic in nature. Moran seems to emphasize the former, arguing that intentional action depends formally on the intentions in light of which an agent acts. Velleman, by contrast, emphasizes that an agent acts intentionally *if* he – implicitly – understands his action in terms of his mental states which cause it. Setiya resists Velleman's thetic approach to non-observational knowledge, arguing that '[...] intention involves belief, and intentional action provides self-knowledge by making true the content of the intention by which it is motivated.' (Setiya 2003, 373).

Davidson argues that an agent need not know what he is doing intentionally. In his view, an action is intentional if caused by a pair of belief-desires or intention which rationalizes that action (Davidson 2001, 50; Davidson 2006b). Among such actions, there are some which an agent does not knowingly perform. For example, someone making carbon copies may intentionally produce ten copies in one cycle, while being quite sceptical that he can produce them in one cycle. This would be the case, for example, if his *desire* to make ten carbon copies and his *belief* that he must press hard in order to make ten carbon copies in one cycle *cause* him to try to make ten carbon copies in one cycle. If this attempt is successful, then he would have produced ten carbon copies intentionally. While acting, however, he did not know that he was making ten carbon copies intentionally.

In critical response to Davidson, Setiya argues that the person making carbon copies at least knows that he is performing the *basic actions* required to make ten carbon copies in one cycle (Setiya 2003, 363). For example, he knows that he is pushing hard in order to make ten carbon copies in one cycle. Setiya, in turn, is criticized by Paul, who argues that an agent can perform a *basic* action intentionally, despite not knowing that he is performing it (Paul 2009, 8; cf. Setiya 2011). For example, a man with a temporarily paralyzed arm may try to raise his arm and be quite sceptical about the proposition that he can raise it. Yet if his attempt is successful, then he is raising it intentionally.

It has been discussed that the non-observational knowledge thesis merely implies that an agent knows what he is *doing* intentionally, not what he *does* intentionally (Falvey 2000, 21–2; M. Thompson 2011; cf. M. Thompson 2008 chapter 8; Kenny 2003 chapter 8). On this reading, the thesis would merely require, for example, that the person making ten carbon copies intentionally knows that he is intentionally making ten carbon copies in one cycle, irrespective of whether or not he makes them intentionally. Or it would require that the paralyzed patient knows that he is raising his arm, irrespective of whether or not he raises his arm. Neither Davidson's nor Paul's argument considers this aspect of the non-observational knowledge thesis explicitly. However, as Falvey mentions, an agent may non-observationally know his intention to do something, but not that he is doing it.

> The openness of the progressive allows for interruptions of actions-in-progress, including changes of mind. But from the fact that an event or process of a given type could not have been completed in the circumstances, it does

seem to follow that no event of that type could have been underway. (Falvey 2000, 24)

What an agent is doing intentionally depends on what he could have completed. An agent may not know that he cannot complete the course of action which he intends to perform. In this sense, then, although an agent may have an intention to do something and know his intention, he does not know that he is doing that action.

To maintain the non-observational knowledge thesis, we might argue that such knowledge is a *blue-print for* rather than a *representation of* one's doing it. That is to say, an agent may know that he is doing something, even if, in Davidson's sense, he does not know what he does *nor*, in the sense discussed by Falvey, that he is doing it. Such knowledge would not be factive, but would be knowledge based on which he is acting.

Alternatively, we may deny that non-observational knowledge is essential for intentional action. This would leave us the task of explaining how it is that an agent generally has non-observational knowledge of his action. Paul adopts this latter strategy (Paul 2009; Paul 2011; cf. Grice 1973). On her account, an agent non-observationally knows his intentional action because (i) he non-observationally knows his intentions and (ii) he knows from past experience that, generally, he does what he intends to do. Based on this, an agent infers – independently of observing it – that he is doing what he knows he intends to do. In addition to the sceptical worry about whether he can legitimately infer from past experiences the general thesis that he does what he intends, there is the question of how it is that an agent non-observationally *knows* what he intends to do. In answer to the second question, Paul argues that – by experience – an agent knows that he forms an intention to do something if he decided to perform that action. Hence, if he decided to perform a particular action, then he can safely assume that he intends to perform that action.

The above-mentioned accounts of non-observational knowledge are incompatible with the thesis that agential self-consciousness is independent of a capacity to consciously act for a reason. That is to say, irrespective of whether or not they explain non-observational knowledge as thetic or telic knowledge, and irrespective of whether or not they explain it as a condition of intentional action, they would *all* entail that agential self-consciousness depends on a capacity to consciously act for a reason. This is obvious with regard to the practical-knowledge approach, given that this approach would have

to explain agential self-consciousness as a form of practical consciousness. The theoretical knowledge approach also accounts for self-knowledge as the same sort of mental entity, i.e., belief, in terms of which it explains action.

The epistemic approach, however, offers better prospects for the extraordinary thesis about agential self-consciousness than the practical-knowledge approach. In particular, Paul's model of the relation between non-observational knowledge and action seems helpful because it explains an agent's consciousness of his action as something which is a *by-product of* rather than a *condition for* intentional action. The basic problem with Paul's account, however, is that it explains an agent's non-observational knowledge on the assumption that his actions are caused by mental states.

Consequently, in order to develop, by reference to models of knowledge, the thesis that agential self-consciousness is independent of a capacity to consciously act for a reason, we must assume that such knowledge is epistemic, i.e., not practical. What is more, we must explain it as a by-product rather than a condition of action. However, in contrast to Paul's approach, we must explain how an agent can non-observationally know his rational action which is not caused by his mental states. I will explain this in the final two sections of the present chapter, but first I will contrast the extraordinary thesis about agential self-consciousness with accounts of self-consciousness in general.

6.2 Self-consciousness – Traditional Accounts

Traditionally, philosophers explain self-consciousness as 'reflexive consciousness' – as if it were wholly a feature of consciousness states or faculties. The general idea is that self-consciousness is an integral feature of mental states, *either* in the sense that a mental state is marked by consciousness of that mental state *or* because it depends on the reflexive consciousness of a subject who has that mental state. This view of self-consciousness has invited theories of 'the self', according to which the self is merely a bundle of perceptions or, in contrast, a unity in consciousness which underlies these mental states (cf. Patten 1976). Irrespective of whether or not there is a self beyond mental states, on this view self-consciousness would differ from object consciousness as the former is (merely) a mind's consciousness of *itself*, whereas the latter is not. How is it that such consciousness is possible?

To explain this, we could assume that mental states are internally transparent or that a subject who has these mental states has an 'inner sense' or 'reflective capacities' by

which he conceives his mental states. However, if we did, then we would have to explain how a subject manages to *identify* himself and *attribute* mental states to himself. What is more, we would have to explain the fact that he conceives himself as the object and the subject of his inner sense or reflective capacities.

Alternatively, we could assume that self-consciousness is a constitutive feature of consciousness. On this model, self-consciousness would be a pre-condition for having mental states at all. In other words, there would be no mental states except insofar as they belong to a subject who has identified himself and adopted these mental states. If we take this approach, then we would have to explain the possibility of such reflexive consciousness activity in order to assume a subject with mental states. To explain the possibility of such reflexive consciousness, Fichte emphasized that *I is spontaneously reflexive conscious activity*. However, as Tugendhat emphasizes, this explains self-consciousness as a creation ex nihilo. Fichte, in a later work, explained reflexive conscious activity as an effect of rational demands on such activity made by other rational agents. In light of their demands, the conscious agent will conceive himself as someone who has to act rationally (Baur 2000; cf. Neuhouser 1990). Hegel pursued a similar line of argument, arguing that self-consciousness is possible only among mutually recognizing conscious agents (cf. Brandom 2007) .

In Tugendhat's view, self-consciousness does not reduce to subjective self-identification. There is no need to explain self-consciousness in terms of a mind's capacity to perceive and identify itself, nor to explain it as a capacity for reflexive conscious self-constitution. I will explain this view, and in particular its implications for the extraordinary thesis about agential self-consciousness, in the remaining two sections of this chapter.

6.3 Tugendhat on Epistemic Self-consciousness

Tugendhat makes a distinction between what he calls 'mediate' and 'immediate' self-consciousness (Tugendhat 1979, 27).[8] In contrast to *immediate* self-consciousness, a

---

[8] Tugendhat claims that human beings relate consciously to themselves in a practical and theoretical way. We are self-determining and self-knowing persons. I focus on his account of self-knowledge because, as I argued earlier in this chapter, I need a thetic account of self-consciousness. Given that Tugendhat's accounts for self-knowledge and self-determination are relatively independent of each other, I suppose I can proceed in the way I do.

person's *mediate* self-consciousness springs from self-observation or from inferences he draws about himself based on observations he makes. Mediate self-consciousness, for example, could be his consciousness that he is now wearing a red sweater or his consciousness that he wore a red sweater as a baby.

Tugendhat argues that self-consciousness, or the relevant form of it, is the sort of *immediate* consciousness that one has of one's *psychological* states. For example, someone who is in pain need not observe himself or draw inferences about himself to tell that he has pain. He immediately *conceives* that he is in pain. What is more, he does so *infallibly*. Hence, Tugendhat emphasizes that an account of immediate self-consciousness would need to be an account of *immediate* self-*knowledge*. It would need to explain that a self-conscious person has *immediate* knowledge of his psychological condition. Hence, self-consciousness is best analysed as a 'I know that I psi.' While uttering such a proposition, a person applies immediately and infallibly two predicates to 'I'. He claims that 'I' 'is in *psi*' and that 'I' '*knows* that it is in psi'.

For this to be possible, a person needs no capacity to identify himself. Instead, he might merely have a capacity to *indicate* himself with 'I' as – among other things – an entity which is identifiable in a linguistic community as a member of it. 'I' is a referring expression but not a concept or a name; it is merely a word by which a language-user indicates itself from a communal standpoint (cf. Anscombe 1975). So in contrast to traditional approaches, Tugendhat emphasizes that 'I' refers first and foremost to an entity which is an identifiable member of a linguistic community, i.e., not primarily a subject of consciousness.

However, he might be a member with mental states. Suppose that – in addition to a capacity to indicate himself – he has a capacity to apply mental-state predicates to himself. If so, then he would not only be capable of indicating himself as someone who is identifiable for other language-users, but also as someone who other language-users would agree has specific mental states.

On Tugendhat's account, it is only because a person and his mental states are *public* that a language-user has a capacity to linguistically indicate himself *and* predicate mental states to himself. Tugendhat illustrates his account with the example of pain. A language-user's capacity to say, 'I am in pain' depends on two facts: that he is a member of a linguistic community *and* that he shows pain behaviour *iff* he is in pain. These two

facts make it possible for other members of the community to identify him and to tell whether or not he is in pain. Within this community, a language-user's use of 'I' is that by which he indicates himself. Hence, a language-user cannot use 'I' and misindicate himself, even though he cannot identify himself. His 'I am in pain' would be true *if* 'he is in pain' is true of him. The use of 'he' and 'I' depend on one and the same *rule*: that the member of a linguistic community uses 'I' to indicate himself as a member whom other members can identify and *would* indicate 'he'. The use of the 'pain' predicate, in turn, depends on a *rule* that the pain predicate is applied to those who show pain behaviour.

Using 'I' or applying the 'pain' predicate to oneself is independent of self-identification *or* self-observation. Hence, it cannot explicitly depend on the rules for language use which I mentioned above. In other words, the use *merely* implicitly depends on these rules. All that a language-user needs to attribute pain states to himself is a capacity to use 'I' and predicate 'pain' to himself, in accordance with these public rules. A person's capacity to indicate himself and predicate mental states to himself may merely depend on a linguistic capacity which is – and continues to be – conditioned so that its exercise *accords* with public language rules.

On my reading of it, Tugendhat offers an account which allows us to challenge two assumptions which we would ordinarily make about the capacity to indicate oneself and predicate states to oneself *and which stand in the way of developing the extraordinary account of agential self-consciousness.* On the one hand, we can ask whether this capacity should be (partially) conscious. By implication we can ask whether self-consciousness, conceived in terms of this capacity, depends on a capacity to consciously act. On the other hand, we can ask whether the predicated states need to be mental ones. For example, the question might be whether an agent can exercise this capacity to ascribe to himself non-conscious motivational responsiveness to reasons.

In terms of the leading question of my dissertation, these two questions are vital for developing the extraordinary thesis about agential self-consciousness. For if we developed a negative answer to both, we could try to explain agential self-consciousness in terms of the exercise of a capacity to make explicit that one is motivationally responsive to reasons.

6.4 Mere Agential Self-consciousness

> Was is der natürliche Ausdruck einer Absicht? – Sieh eine Katze an, wenn sie sich an einen Vogel heranschleicht; oder ein Tier, wenn es entfliehen will
>
> ((Verbindung mit Sätzen über Empfindungen))
>
> (Wittgenstein 2003, para. 647)

Wittgenstein suggested that an intention is similar to pain, in the sense that both have a natural expression. We can debate whether or not Wittgenstein would agree that intentions are mental states like pain states are, or, generally, whether or not we must conceive intentions as mental states. Relying on the mind-independent, generous account of rational agency, I hope to circumvent talk of intentions. I emphasize that we might conceive the cat's behaviour, at least, as a natural expression of his motivational responsiveness to reasons. In addition, I emphasize that there is no need to assume that such motivational responsiveness depends on the mind.

Imagine an agent capable of acting for reasons *and* capable of indicating himself and predicating that he is motivationally responsive to certain reasons. Such an agent would have a capacity to make explicit that he is acting for a reason. Just as pain and pain behaviour are interdependent, so are motivational responsiveness to reasons and rational action. In other words, motivational responsiveness to reasons is interdependent with an action which is performed *because* that action is a factor for something else. The question of whether or not an action is performed because it is a factor for something else is similar to the question of whether or not something is pain behaviour. To answer the former question, one needs to distinguish *merely apparent* from *real* rational action. Distinguishing these forms of actions from each other is a matter of determining the extent to which an agent's behaviour would change if the facts about what his action is a factor of changed.

If motivational responsiveness to reasons has a natural expression, like pain has, then we could explain someone's capacity to make explicit that he is acting for a reason as being on a par with someone's capacity to make explicit that he is in pain. Hence, an agent might have a capacity to indicate himself and predicate states to himself *irrespective* of whether or not these states are mental. What is more, his capacity to do so need not be a conscious one. If not, then a rational agent would have a capacity to make explicit that he acts for a reason *independently of a capacity to consciously act for a reason.*

This brings us quite close to the extraordinary thesis about agential self-consciousness. However, there is one more obstacle which we must face. If the capacity to act for a reason and the capacity to make explicit that one is acting for a reason are independent of consciousness, then why *should* we ascribe agential *self-consciousness* to an agent who has both? I do not stress that we *should*. Rather, I would imply that *if* we already ascribe agential self-consciousness, then we might as well explain it as an effect that these two capacities have on consciousness. Let me elaborate on this in what follows.

We may attribute to an agent a capacity to indicate himself and predicate non-mental practical states to himself in *accordance* with the rules of public language. Like his capacity to predicate mental states to himself in accordance with rules of public language, this capacity is independent of a capacity for self-identification or self-observation. That is to say, he need not identify himself nor observe that he is acting for a reason; instead, he need only exercise a linguistic capacity which is conditioned so that its exercise yields a linguistic expression of the publicly identifiable fact that the agent is acting for a reason. In the first part of this dissertation, I have developed the view that the capacity to act for a reason is mind-independent. In the present and previous chapters, I have been developing the thesis that the capacity to make explicit that one is acting for a reason is mind-independent. This approach suggests a radical separation between these two capacities, on the one hand, and consciousness, on the other hand. How could an appeal to these two capacities explain agential self-consciousness?

Based on my reading of Tugendhat's account, we could explain that an agent can, independent of observation, justifiably *say*, 'I am motivationally responsive to a reason.' Explaining this, however, falls short of explaining agential self-consciousness, as the latter would need to explain that an agent can, independently of observation, justifiably *say*, 'I *know* that I am motivationally responsive to a reason.'

If *knowledge* had a natural expression by virtue of which it is public, then we could explain an agent's saying, 'I know that I am motivationally responsive to a reason' on a par with his saying, 'I am in pain' or 'I am motivationally responsive to a reason.' Does *knowledge* have a natural expression?

It has been argued by Stout, in a neo-behaviourist vein, that mental states and behavioural dispositions are – in principle – two sides of the same coin (Stout 2006). In

contrast to traditional behaviourism, Stout argues that mental states do not reduce to behaviour dispositions. Rather, he suggests that an agent believes that it is raining *only if disposed to act as if it is raining.* Stout suggests that a belief that something is the case equals a disposition to act on the assumption that it is the case. Consequently, Stout's account might be taken to imply that the *belief* that something is the case has as its natural expression *acting* on the assumption that it is the case. Although such an account is plausible in its own terms, it would entail the sort of mind-dependent account of action which I cannot adopt.

There is no need to proceed as if a belief that such and such is so has acting as if such and such is so as its natural expression. On the one hand, the generous account of rational action explains that actions are independent of beliefs. So actions need not be signs of mental states at all. On the other hand, propositional attitudes such as beliefs may be identifiable independently of actions regulated by these beliefs, e.g., by the *effects* that the exercise of a capacity to express propositions *has on the mind.* For example, whether or not this capacity has such an effect might depend on whether or not its *exercise* provokes mental states, such as *feelings*, *emotions* or *moods* concerning the facts being expressed, which (i.e., the feelings, emotions or moods) have natural expressions. If it does, then the mental states provoked by this exercise can be explained as effects of beliefs which, in turn, are to be explained in terms of the exercise of the linguistic capacity to express facts.

On this model, beliefs have feelings, emotions or moods as their natural expressions, and these feelings, emotions or moods have behaviour as their natural expressions. Generally speaking, feelings, emotions and moods might involve, but are independent of, beliefs. That is to say, agents may be in pain or be restless or angry even if they lack a capacity to express facts linguistically and, consequently, lack a capacity for belief. This might be the case, for example, with a cat that gets excited upon seeing a bird high up in a tree. He may not believe that there is a bird high up in the tree, yet he may be excited by the fact that there is a bird high up in the tree. We should take these mental states as natural expressions of beliefs only if there is reason to assume that they are. On my view, a reason for assuming this would be if the state is aroused by the exercise of a capacity to express facts. An agent who is mentally affected by his capacity to make explicit that he is acting for a reason can be said to believe that he is acting for a reason.

If belief, like pain, has – in this latter way – a natural expression, then an agent's capacity to say, 'I believe that I am motivationally responsive' could be explained on a par with his capacity to say, 'I am in pain.' The former capacity, like the latter, could be independent of a capacity for self-identification or self-observation. Nevertheless, it could be conditioned so that its exercise gives linguistic expression to the practical and mental condition of the agent who has it. What is more, this account of beliefs is compatible with the extraordinary thesis about agential self-consciousness. It allows us to ascribe to a rational agent a capacity for agential self-consciousness that is independent of a capacity to consciously act for a reason.

On this account of it, agential self-consciousness is a feature of an agent who has (i) a capacity to act for a reason, (ii) a capacity to make explicit that he is acting for a reason and (iii) a mind. The former two are independent of the mind. The first is independent of the second. The second depends on the former. The mind depends on the second – and consequently on the former – only if it involves beliefs which are effects of exercising the second capacity. Such an agent would have agential self-consciousness independently of a capacity to consciously act for a reason. Of course, he might also have a capacity to consciously act for a reason, but if so then only due to still further capacities which are dispensable for the former two.

For example, our coffee-drinking agent may have a capacity to act for a reason, a capacity to make explicit that he is acting for a reason *and* a mind. His capacity to drink coffee and his capacity to make explicit that he is drinking coffee to stay alert might be mind-independent. Of course, his capacity to make explicit that he is drinking coffee to stay alert depends on his capacity to drink coffee to stay alert. However, his mind would depend on his capacity to make explicit that he is drinking coffee to stay alert – and consequently on his capacity to drink coffee to stay alert – only if it involves the belief that he is drinking coffee to stay alert. In this sense, our coffee-drinking agent would have agential self-consciousness independently of a capacity to consciously act for a reason.

6.5 Conclusion

In this chapter I have completed my attempt to develop the thesis that agential self-consciousness is independent of a capacity to consciously act for a reason. I have positioned my thesis against *two* ideas about self-consciousness. On the first, an agent's consciousness that he is acting intentionally is a condition of his acting intentionally *or*

a result of mental causes of action. Against this idea, I stressed that there is no reason to assume that rational action depends on consciousness at all. On the second idea, self-consciousness depends on subjective self-identification – and resultantly on the assumption that it is consciousness of mental states or a mind. I have introduced Tugendhat's account of self-consciousness to argue that self-consciousness is independent of subjective self-identification and requires merely a capacity *to indicate* oneself as an entity which can be identified by others *and* to predicate mental states to oneself. From this account, I have developed the claim that an agent can have non-observational knowledge that he is acting for a reason independently of a capacity to consciously act for a reason.

7 – General Conclusion

I opened my dissertation with two observations: that we rationalize the behaviour of – presumably – unconscious agents, and that we sometimes do not do and even feel powerless – to do what we think would be factors for a greater good. I argued that these phenomena are somewhat incompatible with our ordinary self-concepts, given that we typically assume that our agential self-consciousness marks our capacity to act rationally, i.e., to act consciously for a reason. I emphasized that these phenomena, and in particular our understanding of them in terms of our ordinary self-concepts, are part and parcel of our idea that agential self-consciousness is a sufficient condition for praise or blame, our understanding of normativity in general, and the type of moral distinctions which we make between self-conscious human agents, on the one hand, and other life forms, on the other.

The extraordinary thesis states that agential self-consciousness is independent of a capacity to consciously act for a reason. If true, this thesis would drastically affect our

understanding of rational agency, and in particular our denial of such agency to agents who presumably lack agential self-consciousness. What is more, it would require us to explain whether, and if so why and to what extent, our agential self-consciousness depends on a capacity to consciously act for a reason. Hence, in determining whether or not animals or young agents are rational agents, we should not start with the assumption that agential self-consciousness is a defining feature of rational agency, not even in those rational agents who have it.

Moreover, if the extraordinary thesis were true it would change the way we speak of and conceive ourselves when we do not do what we assent would be a factor for a greater good than what we are actually doing. We would need to be careful in assuming that acting this way is a form of irrationally, i.e., a display of defective rational capacities or of a defective exercise of these capacities. We must be careful not to assume that when acting in these way, we rational agents are under the spell of some evil force, are psychologically impaired *or* are compelled by our desires. Also, we must be careful not to proceed as if although we are not doing, we should be and are capable of doing what we assent would be a factor for a greater good. In contrast to making one of these assumptions, we might assume that not doing what we assent would be a factor for a greater good is what we can expect from human beings who have learned to say whether or not their behaviour is a factor for a good or bad thing *quite* independently of learning to do what is a factor for a good and avoiding what is a factor for a bad thing. If the extraordinary thesis were true, this is what we could expect in societies where *learning* is a matter of *being told*, if being told would be insufficient for acquiring a capacity to do as we are told. For in these cases we would not be motivationally responsive to the fact that an action is a factor for the good, yet we might have a capacity to make explicit that we are not motivationally responsive to that fact.

Finally, if true, the extraordinary thesis would facilitate a background thesis for challenging the stories in terms of which we rationalize the ways in which we interact with those who lack a capacity to consciously act for a reason. In particular, this would be relevant to the stories we tell to prioritize the life and projects of conscious agents above the lives and courses of action of unconscious animals or unconscious human beings. In effect, it would form a background against which we can challenge those who mean to rationalize – in terms of the importance of their own projects, such as having cheap meat or dairy products or having a new smartphone – killing and exploiting

animals and destroying or depleting their habitats or natural resources which future generations might need.

I argued that the extraordinary thesis about agential self-consciousness makes sense, once we assume the following:

(i)   that reasons are a specific sort of facts about an action and external to the mind;

(ii)   that acting for reasons, thus conceived, might merely involve mind-independent motivational responsiveness; and

(iii)   that an agent's consciousness that he is acting for a reason might merely be an effect on his consciousness of his non-mental (rudimentary) linguistic capacity to make explicit that he is motivationally responsive to a reason.

I contended that assuming the first two make sense in light of Anscombe's definition of intentional action, Taylor's account of behaviour explanation and in light of externalism about reasons. Furthermore, I contended that the third makes sense in light of Wittgenstein's claim that mental states cannot be private, Mead's account of meaning and significant gestures and Tugendhat's Wittgensteinian account of epistemic self-consciousness.

I have *neither* argued that the extraordinary thesis about agential self-consciousness is true *nor* that it coheres with contemporary philosophy. By implication, I have not argued for what would follow if the thesis were true.

However, I have argued that this thesis can be *developed* from certain assumptions about rational agency and self-consciousness. That is, the thesis makes sense as an implication of these assumptions once they are combined. On my view of it, the fact that we can develop a thesis with such implications indicates that we cannot take the risk of *assuming* that the extraordinary thesis cannot be true. In other words, I take my dissertation to imply that we must discuss the extraordinary thesis more extensively – in terms of its possibility, coherence and truth value.

We might find that, as a special case of agential self-consciousness, *our* agential self-consciousness *marks* a capacity to consciously act for a reason. But if we find that our agential self-consciousness does not mark a capacity to consciously act for a reason, we would have to find new ways to mark ourselves in relation to the rest of living nature.

But most fundamentally, I would take this thesis as another occasion to scrutinize the basic self-concept which supports the gap between the social order and the living order in which we partake. The extraordinary thesis would provide for a self-concept which might to some extent bridge the gap between us and the order of unconscious agents, on the one hand, and between us and the apparant 'irrational' aspects of life, on the other.

Literature

Adorno, Theodor W., and Max Horkheimer. 1989. *Dialektik Der Aufklarung*. Frankfurt
     am Main: Fischer Taschenbuch Verlag GmbH.

Agosta, Lou. 2011. "Empathy and Sympathy in Ethics." *Internet Encyclopedia of
     Philosophy*. http://www.iep.utm.edu/emp-symp/.

Allen, Colin. 2011. "Animal Consciousness." In *The Stanford Encyclopedia of Philosophy*,
     ed. Edward N. Zalta. Winter 2011.
     http://plato.stanford.edu/archives/win2011/entries/consciousness-animal/.

Alvarez, Maria. 2009. "How Many Kinds of Reasons?" *Philosophical Explorations* 12 (2):
     181–193.

———. 2010. *Kinds of Reasons : an Essay in the Philosophy of Action*. Oxford ;New York:
     Oxford University Press.

Andrews, Kristin. 2011. "Animal Cognition." In *The Stanford Encyclopedia of Philosophy*,
     ed. Edward N. Zalta. Summer 2011.
     http://plato.stanford.edu/archives/sum2011/entries/cognition-animal/.

Anon. 2007. *Addicted Monkey*.
    http://www.youtube.com/watch?v=uR1yAaJoZbE&feature=youtube_gdata_pla
    yer.

———. 2008a. *Wild African Vulture Birds Scavage Bones of Dead Animals - BBC Wildlife*.
    http://www.youtube.com/watch?v=zxj9YO4Qtx0&feature=youtube_gdata_pla
    yer.

———. 2008b. *Alex The Parrot*.
    http://www.youtube.com/watch?v=ldYkFdu5FJk&feature=youtube_gdata_pla
    yer.

Anscombe, G. E. M. 1956. "Intention." *Proceedings of the Aristotelian Society* 57: 321–
    332.

———. 1958. "Modern Moral Philosophy." *Philosophy* 33 (124) (January 1): 1–19.

———. 1975. "The First Person." In *Mind and Language*, ed. Samuel Guttenplan, 45–65.
    Oxford University Press.

———. 1979. "Under a Description." *Noûs ,* 13 (2): 219–233.

———. 2000. *Intention*. 2nd ed. Cambridge Mass.: Harvard University Press.

Apel, Karl-Otto. 1975. "The Problem of Philosophical Fundamental-grounding in Light
    of a Transcendental Pragmatic of Language." *Man and World* 8 (3): 239–275.

Balaban, Oded. 1990. "Praxis and Poesis in Aristotle's Practical Philosophy." *The Journal
    of Value Inquiry* 24 (3): 185–198.

Baur, Michael. 2000. *Fichte : Foundations of Natural Right*. Cambridge: Cambridge
    University Press.

Bayne, Tim. Forthcoming. "Agency as a Marker of Consciousness." In *Decomposing the
    Will*, ed. T Vierkant, J Kiverstein, and A Clark. Oxford: Oxford University Press.

———. 2011. "The Sense of Agency." In *The senses : classic and contemporary
    philosophical perspectives*, ed. Fiona Macpherson. New York: Oxford University
    Press.

Bayne, Tim, and Elisabeth Pacherie. 2007. "Narrators and Comparators: The
    Architecture of Agentive Self-awareness." *Synthese* 159 (3) (September 28):
    475–491.

Beauchamp, Tom L. 1999. "The Failure of Theories of Personhood." *Kennedy Institute of
    Ethics Journal* 9 (4) (December): 309–324.

Bengson, John, and Marc A. Moffett, eds. 2012. *Knowing How: Essays on Knowledge,
    Mind, and Action*. Oxford University Press, USA.

Bermúdez, José Luis. 2007. "Thinking Without Words: An Overview for Animal Ethics." *The Journal of Ethics* 11 (3): 319–335.

Beyleveld, Deryck. 1991. *The Dialectical Necessity of Morality : an Analysis and Defense of Alan Geworth's Argument to the Principle of Generic Consistency*. Chicago u.a: Univ. of Chicago Press.

Beyleveld, Deryck, and Gerhard Bos. 2009. "The Foundational Role of the Principle of Instrumental Reason in Gewirth's Argument for the Principle of Generic Consistency: A Response to Andrew Chitty." *Kings College Law Journal* 20 (1): 1–20.

Bittner, Rüdiger. 2001. *Doing Things for Reasons*. Oxford: Oxford University Press.

Bortolotti, Lisa. 2007. "Disputes over Moral Status: Philosophy and Science in the Future of Bioethics." *Health Care Analysis: HCA: Journal of Health Philosophy and Policy* 15 (2) (June): 153–158.

Boyle, Deborah. 2003. "Hume on Animal Reason." *Hume Studies* 29 (1): 3–28.

Brandom, Robert B. 2007. "The Structure of Desire and Recognition." *Philosophy & Social Criticism* 33 (1) (January 1): 127 –150. doi:10.1177/0191453707071389.

Bratman, Michael. 1987. *Intention, Plans, and Practical Reason*. Cambridge Mass.: Harvard University Press.

Broome, John. 1997. "Reasons and Motivation: John Broome." *Aristotelian Society Supplementary Volume* 71 (1) (June): 131–146.

———. 1999. "Normative Requirements." *Ratio* 12 (4): 398–419.

———. 2002. "Practical Reasoning." *Reason and Nature: Essays in the Theory of Rationality*: 85.

———. 2005. "Does Rationality Give Us Reasons?" *Philosophical Issues* 15: 321–337.

———. 2007. "Is Rationality Normative?" *Disputatio* 2 (23): 161–178.

Candlish, Stewart, and George Wrisley. 2012. "Private Language." In *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta. Summer 2012. http://plato.stanford.edu/archives/sum2012/entries/private-language/.

Carruthers, Peter. 2011. *The Opacity of Mind: An Integrative Theory of Self-Knowledge*. Oxford University Press, USA.

Casati, Roberto, and Achille Varzi. 2010. "Events." In *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta. Spring 2010. http://plato.stanford.edu/archives/spr2010/entries/events/.

Cassirer, Ernst, and Maureen Lukay. 2006. *An essay on man : an introduction to a philosophy of human culture*. Hamburg: F. Meiner.

Chisholm, Roderick. 1976. "The Agent as Cause." In *Action Theory: Proceedings of the Winnipeg Conference on Human Action, Held at Winnipeg, Manitoba, Canadá, 9-11 May 1975*, ed. Myles Brand, 199–211. Dordrecht; Boston: Springer Science & Business.

Colapietro, V. 2006. "Toward a Pragmatic Conception of Practical Identity." *Transactions of the Charles S. Peirce Society* 42 (2): 173–205.

Couch, Mark B. 2011. "Causal Role Theories of Functional Explanation." *Internet Encyclopedia of Philosophy*. http://www.iep.utm.edu/func-exp/.

Crowell, Steven. 2010. "Existentialism." In *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta. Winter 2010. http://plato.stanford.edu/archives/win2010/entries/existentialism/.

Davidson, Donald. 1970. "How Is Weakness of the Will Possible?" In *Moral Concepts*, ed. Joel Feinberg. Oxford University Press.

———. 1982. "Rational Animals." *Dialectica* 36 (4) (December 1): 317–327.

———. 2001. "Agency." In *Essays on Actions and Events*, 43–58. Oxford: Oxford University Press.

———. 2006a. "Actions, Reasons, and Causes." In *The Essential Davidson*, 24–36. Oxford: OUP.

———. 2006b. "Intending." In *The Essential Davidson*, 122–37. Oxford: OUP.

DeGrazia, D. 1999. "Animal Ethics Around the Turn of the Twenty-first Century." *Journal of Agricultural & Environmental Ethics* 11: 111–129.

Dennett, Daniel C. 1989. *The Intentional Stance*. reprint. A Bradford Book.

Dillon, Robin S. 1997. "Self-Respect: Moral, Emotional, Political." *Ethics* 107 (2) (January 1): 226–249.

Ehring, Douglas. 1985. "Dispositions and Functions: Cummins on Functional Analysis." *Erkenntnis (1975-)* 23 (3) (November 1): 243–249.

Engelhardt Jr, H.T. 1988. "Medicine and the Concept of Person." In *What Is a Person*, ed. Michael F. Goodman, 169–82.

Eshleman, Andrew. 2009. "Moral Responsibility." In *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta. Winter 2009. http://plato.stanford.edu/archives/win2009/entries/moral-responsibility/.

Falvey, Kevin. 2000. "Knowledge in Intention." *Philosophical Studies* 99 (1): 21–44.

Foot, Philippa. 2001. *Natural Goodness*. Oxford University Press.

Frank, Manfred. 2002. "Self–consciousness and Self–knowledge: On Some Difficulties with the Reduction of Subjectivity." *Constellations* 9 (3) (September 1): 390–408.

Frankfurt, Harry G. 1971. "Freedom of the Will and the Concept of a Person." *The Journal of Philosophy* 68 (1): 5–20.

———. 1978. "The Problem of Action." *American Philosophical Quarterly* 15 (2) (April 1): 157–162.

———. 2004. *The reasons of love*. Princeton, N.J.: Princeton University Press.

Gallagher, Shaun. 1996. "The Moral Significance of Primitive Self-Consciousness: A Response to Bermúdez." *Ethics* 107 (1) (October 1): 129–140.

———. 2007. "The Natural Philosophy of Agency." *Philosophy Compass* 2 (2): 347–357.

Gallagher, Shaun, and Dan Zahavi. 2010. "Phenomenological Approaches to Self-Consciousness." In *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta. Winter 2010. http://plato.stanford.edu/archives/win2010/entries/self-consciousness-phenomenological/.

Gauthier, David. 1987. *Morals By Agreement*. Oxford University Press, USA.

Gert, Bernard. 1996. "Hobbes's Psychology." In *The Cambridge Companion to Hobbes*, ed. Tom Sorell, 157–174. Cambridge: Cambridge University Press.

Gertler, Brie. 2011a. "Self-Knowledge." In *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta. Spring 2011. http://plato.stanford.edu/archives/spr2011/entries/self-knowledge/.

———. 2011b. *Self-Knowledge*. New York: Taylor & Francis.

Graham, George. 2010. "Behaviorism." In *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta. Fall 2010. http://plato.stanford.edu/archives/fall2010/entries/behaviorism/.

Grice, H. 1973. "Intention and Uncertainty." *Proceedings of the British Academy*: 263–279.

Gruen, Lori. 2010. "The Moral Status of Animals." In *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta. Fall 2010. http://plato.stanford.edu/archives/fall2010/entries/moral-animal/.

Gustafson, Don. 2007. "Neurosciences of Action and Noncausal Theories." *Philosophical Psychology* 20 (3): 367–374.

Gustafsson, M. "Anscombe's Bird, Wittgenstein's Cat." *Unpublished*. https://www.abo.fi/sitebuilder/media/6840/anscombesbirdwittgensteinscatforskis.pdf.

Haack, Robin. 1982. "Wittgenstein's Pragmatism." *American Philosophical Quarterly* 19 (2) (April 1): 163–171.

Habermas, Jürgen. 1988a. "'Individuierung Durch Vergesellschaftung: Zu G. H. Meads Theorie Der Subjektivität.'" In *Nachmetaphysisches Denken: Philosophische Aufsätze*, P. 187 ff. Frankfurt a. M.: Suhrkamp.

———. 1988b. "Morality and Ethical Life: Does Hegel's Critique of Kant Apply to Discourse Ethics." *Northwestern University Law Review* 83: 38.

Heyes, Cecilia, and Anthony Dickinson. 2007. "The Intentionality of Animal Action." *Mind & Language* 5 (1): 87–103.

Homiak, Marcia L. 2000. "Hume's Ethics: Ancient or Modern?" *Pacific Philosophical Quarterly* 81 (3): 215–236.

Hornsby, Jennifer. 2004. "Agency and Actions." In *Agency and Action*, ed. Helen Steward and J Hyman. Cambridge: Cambridge University Press.

Humberstone, I. L. 1992. "Direction of Fit." *Mind* 101 (401): 59–83.

Hume, David. 1978. *A Treatise of Human Nature*. Ed. L. A. Selby-Bigge and P. H. Nidditch. Oxford; New York: Clarendon Press ; Oxford University Press.

Hurley, Susan. 2003. "Animal Action in the Space of Reasons." *Mind & Language* 18 (3) (April 30): 231–257.

Hursthouse, Rosalind. 1991. "Arational Actions." *The Journal of Philosophy* 88 (2) (February 1): 57–68.

Kalis, Annemarie. 2011. *Failures of Agency: Irrational Behavior and Self-Understanding*. Lanham: Lexington Books.

Kant, Immanuel. 1999. *Grundlegung Zur Metaphysik Der Sitten*. Ed. Dieter Schonecker. Hamburg: F. Meiner.

———. 2003. *Kritik Der Praktischen Vernunft*. Ed. Horst D. Brandt. Hamburg: Meiner.

Kauppinnen, Antti. Forthcoming. "'Sentimentalism'." Ed. hugh LaFolette. *International Encyclopedia of Ethics*. Wiley-Blackwell. http://tcd.academia.edu/AnttiKauppinen/Papers/1309568/Sentimentalism_International_Encyclopedia_of_Ethics_.

Kenny, Anthony John Patrick. 2003. *Action, Emotion and Will*. London; New York: Routledge.

Kleingeld, Pauline. 2010. "Moral Consciousness and the 'Fact of Reason'." In *Kant's "Critique of Practical Reason"*, 55–72. Cambridge Critical Guides. Cambridge University Press.

Korsgaard, Christine M. 1996. *The Sources of Normativity*. Cambridge; New York: Cambridge University Press.

———. 2009. *Self-constitution : Agency, Identity, and Integrity*. Oxford; New York: Oxford University Press.

Kraut, Richard. 2012. "Aristotleś Ethics." In *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta. Spring 2012. http://plato.stanford.edu/archives/spr2012/entries/aristotle-ethics/.

Luckhardt, C. Grant. 1983. "Wittgenstein and Behaviorism." *Synthese* 56 (3): 319–338.

Mackie, J. L. 2005. *Hume's Moral Theory*. London; Boston: Routledge.

Markell, Patchen. 2007. "The potential and the actual." In *Recognition and power : Axel Honneth and the tradition of critical social theory*, ed. Bert van den Brink and David Owen, 100–32. Cambridge; New York: Cambridge University Press.

Markie, Peter. 2012. "Rationalism Vs. Empiricism." In *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta. Summer 2012. http://plato.stanford.edu/archives/sum2012/entries/rationalism-empiricism/.

Matthews, Gareth. 2010. "The Philosophy of Childhood." In *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta. Winter 2010. http://plato.stanford.edu/archives/win2010/entries/childhood/.

McCarty, Richard. 2009. *Kant's Theory of Action*. Oxford University Press.

McCormick, Matt. 2005. "Kant, Immanuel: Metaphysics." *Internet Encyclopedia of Philosophy*. http://www.iep.utm.edu/kantmeta/.

McDowell, John Henry. 1994. *Mind and World: With a New Introduction*. Harvard University Press.

McLaughlin, Peter. 2001. *What Functions Explain : Functional Explanation and Self-reproducing Systems*. Cambridge: Cambridge University Press.

Mead, George Herbert. 1925. "The Genesis of the Self and Social Control." *International Journal of Ethics* 35 (3) (April 1): 251–277.

Mead, George Herbert, and Charles William Morris. 1967. *Mind, Self, and Society : from the Standpoint of a Social Behaviorist*. Chicago: University of Chicago Press.

Mele, Alfred R. 1987. "Intentional Action and Wayward Causal Chains: The Problem of Tertiary Waywardness." *Philosophical Studies* 51 (1): 55–60.

———. 2000. "Goal-Directed Action: Teleological Explanations, Causal Theories, and Deviance." *Nous* 34 (s14) (October): 279–300.

———. 2004. "Discussion - Velleman on Action and Agency." *Philosophical Studies* 121 (3) (December): 249–261.

Metzinger, Thomas. 2011. "The No-self Alternative." In *The Oxford Handbook of the Self*, ed. S. Gallagher, 442–464. Oxford Handbooks Online.

Millgram, Elijah. 2008. "Practical Reason and the Structure of Actions." *Stanford Encyclopedia of Philosophy*. http://www.science.uva.nl/~seop/entries/practical-reason-action/.

van Miltenburg, Niels. 2012. "Practical Knowledge and Foreseen Side Effects." *Journal of Ethics and Social Philosophy*.

Minton, Arthur J. 1975. "Wright and Taylor: Empiricist Teleology." *Philosophy of Science* 42 (3): 299–306.

Mitchell, Sandra D. 1993. "Dispositions or Etiologies? A Comment on Bigelow and Pargetter." *The Journal of Philosophy* 90 (5): 249–259.

Moran, Richard. 2004a. "Précis of 'Authority and Estrangement: An Essay on Self-Knowledge'." *Philosophy and Phenomenological Research* 69 (2): 423–426.

———. 2004b. "Anscombe on 'practical Knowledge'." *Royal Institute of Philosophy Supplement* 55 (-1): 43–68.

Morse, Stephen J. 2007. "The Non-problem of Free Will in Forensic Psychiatry and Psychology." *Behavioral Sciences & the Law.* 25 (2): 203–220.

———. 2008. "Determinism and the Death of Folk Psychology: Two Challenges to Responsibility from Neuroscience." *Minnesota Journal of Law, Science & Technology* 9: 1.

Neuhouser, Frederick. 1990. *Fichte's Theory of Subjectivity*. Cambridge; New York: Cambridge University Press.

Newen, Albert, and Kai Vogeley. 2003. "Self-representation: Searching for a Neural Signature of Self-consciousness." *Consciousness and Cognition* 12 (4) (December): 529–543.

Newson, Ainsley J. 2007. "Personhood and Moral Status": 277–283.

Nietzsche, Friedrich. 1999. *Jenseits Von Gut Und Böse. Zur Genealogie Der Moral. Herausgegeben Von G. Colli Und M. Montinari.* Ed. Mazzino Montinari. N.-A., Nachdr. Deutscher Taschenbuch Verlag.

Noble, Denis. 1967. "Charles Taylor on Teleological Explanation." *Analysis* 27 (3): 96–103.

Norton, David Fate. 1981. "The Myth of 'British Empiricism'." *History of European Ideas* 1 (4): 331–344.

Nussbaum, Martha C. 1988. "Non-Relative Virtues: An Aristotelian Approach." *Midwest Studies In Philosophy* 13 (1): 32–53.

Olafson, Frederick A. 1975. "Consciousness and Intentionality in Heidegger's Thought." *American Philosophical Quarterly* 12 (2) (April 1): 91–103.

Patten, S. C. 1976. "Hume's Bundles, Self-Consciousness and Kant." *Hume Studies* 2.

Paul, Sarah K. 2009. "How We Know What We're Doing." *Info: Ann Arbor, MI: Scholarly Publishing Office, University of Michigan Library* 9 (11).

———. 2011. "How We Know What We Intend." *Philosophical Studies* (May 18).

Pears, David Francis. 1965. "Predicting and deciding." In *Proceedings of the Britisch Academy, 1964.* Vol. 50.

Pearson, Giles. 2012. *Aristotle on desire*. Cambridge; New York: Cambridge University Press.

Pink, Thomas, and M. W. F Stone. 2004. *The will and human action : from antiquity to the present day*. London: New York : Routledge.

Pippin, Robert B. 1987. "Kant on the Spontaneity of Mind." *Canadian Journal of Philosophy* 17 (2): 449–475.

———. 1991. "Idealism and Agency in Kant and Hegel." *The Journal of Philosophy* 88 (10) (October 1): 532–541.

Polanyi, Michael. 1962. *Personal knowledge : towards a post-critical philosophy*. Chicago: University of Chicago Press.

Pollard, Bill. 2005. "The Rationality of Habitual Actions". Conference Paper. http://www.era.lib.ed.ac.uk/handle/1842/1352.

Proust, J. 2003. "How Voluntary Are Minimal Actions?" In *Voluntary Action*, ed. S Maasen, W Prinz, and J Roth. Oxford: Oxford University Press.

Rawls, John. 1955. "Two Concepts of Rules." *The Philosophical Review* 64 (1): 3–32.

Rochat, Philippe. 2011. "What Is It Like to Be a Newborn?" In *The Oxford Handbook of the Self*, by Shaun Gallagher, 1:57–80. 1st ed. Oxford, England: Oxford University Press.

Ryle, Gilbert. 1945. "Knowing How and Knowing That: The Presidential Address." *Proceedings of the Aristotelian Society* 46. New Series (January 1): 1–16.

———. 2008. *The Concept Of Mind*. Kessinger Publishing, LLC.

Schapiro, Tamar. 2001. "Three Conceptions of Action in Moral Theory." *Nous* 35 (1): 93–117.

Schlosser, Markus E. 2007. "Basic Deviance Reconsidered." *Analysis* 67 (3) (July 1): 186–194.

Schmid, Ulla. 2011. "Where Individuals Meet Society: The Collective Dimensions of Self-evaluation and Self-Knowledge." In *Self-evaluation: Affective and Social Grounds of Intentionality*, ed. Keith Lehrer and Anita Konzelmann Ziv, 254–274. Springer.

Schwitzgebel, Eric. 2010. "Introspection." In *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta. Fall 2010. http://plato.stanford.edu/archives/fall2010/entries/introspection/.

Searle, John R. 1991. "Consciousness, Unconsciousness and Intentionality." *Philosophical Issues* 1 (January 1): 45–66.

Sedgwick, Sally S. 1988. "On the Relation of Pure Reason to Content: A Reply to Hegel's Critique of Formalism in Kant's Ethics." *Philosophy and Phenomenological Research* 49 (1): 59–80.

Setiya, Kieran. 2003. "Explaining Action." *Philosophical Review* 112 (3) (July): 339–393.

———. 2004. "Hume on Practical Reason." *Nous* 38: 365–389.

———. 2008. "Practical Knowledge." *Ethics* 118 (3) (April 1): 388–409.

———. 2010. "Intention." http://plato.stanford.edu/entries/intention/.

———. 2011. "Knowledge of Intention." In *Essays on Anscombe's Intention*, ed. Anton Ford, Jennifer Hornsby, and Frederick Stoutland. Harvard University Press.

Shoemaker, Sydney. 1988. "On Knowing One's Own Mind." *Philosophical Perspectives* 2 (January 1): 183–209.

———. 1994. "Self-Knowledge and 'Inner Sense': Lecture III: The Phenomenal Character of Experience." *Philosophy and Phenomenological Research* 54 (2) (June 1): 291–314.

Sie, Maureen, and Arno Wouters. 2009. "The BCN Challenge to Compatibilist Free Will and Personal Responsibility." *Neuroethics* 3 (2) (December 15): 121–133.

Silva, Filipe Carreira Da. 2007. *G.H. Mead: A Critical Introduction*. Polity.

Solomon, Robert C. 1992. "Existentialism, Emotions, and the Cultural Limits of Rationality." *Philosophy East and West* 42 (4) (October 1): 597–621.

Stanley, Jason. 2011. *Know How*. Oxford University Press, USA.

Steward, Helen. 2009. "Animal Agency." *Inquiry* 52 (3): 217.

Stout, Rowland. 1996. *Things That Happen Because They Should : a Teleological Approach to Action*. New York: Oxford University Press.

———. 2005. *Action*. Central Problems of Philosophy. Chesham [U.K.]: Acumen.

———. 2006. *The Inner Life of a Rational Agent : in Defence of Philosophical Behaviourism*. Edinburgh: Edinburgh University Press.

Strawson, P. F. 1959. *Individuals, an Essay in Descriptive Metaphysics*. London: Methuen.

Stroud, Sarah. 2009. "Weakness of Will." *Stanford Encyclopedia of Philosophy* 8 (4).

Sullivan, Denis F. 2009. "Anscombe on Freedom, Animals, and the Ability to Do Otherwise." *Proceedings of the American Catholic Philosophical Association* 81 (February 23): 231–240.

Taylor, Charles. 1964. *The Explanation of Behaviour*. International Library of Philosophy and Scientific Method. London; New York: Routledge & Kegan Paul; Humanities Press.

———. 1967. "Teleological Explanation: A Reply to Denis Noble." *Analysis* 27 (4): 141.

———. 1985. *Human Agency and Language*. Cambridge University Press.

———. 1992. *Sources of the Self: The Making of the Modern Identity*. Harvard University Press.

———. 1993. "Engaged Agency and Background in Heidegger." In *The Cambridge Companion to Heidegger*, ed. Charles B. Guignon, 317–336. Cambridge University Press.

Thompson, Evan. 2007. *Mind in Life: Biology, Phenomenology, and the Sciences of Mind*. Belknap Press.

Thompson, Michael. 2008. *Life and Action : Elementary Structures of Practice and Practical Thought*. Cambridge, Mass.: Harvard University Press.

———. 2011. "Anscombe's Intention and Practical Knowledge." In *Essays on Anscombe's Intention*, ed. Anton Ford, Jennifer Hornsby, and Frederick Stoutland. Harvard University Press.

Timmermann, Jens. 2010. "Reversal or Retreat? Kant's Deductions of Freedom and Morality." In *Kant's "Critique of Practical Reason"*, 73–89. Cambridge Critical Guides. Cambridge University Press.

Tugendhat, Ernst. 1979. *Selbstbewusstsein Und Selbstbestimmung : Sprachanalytische Interpretationen*. Frankfurt am Main: Suhrkamp.

Uzgalis, William. 2010. "John Locke." In *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta. Winter 2010. http://plato.stanford.edu/archives/win2010/entries/locke/.

Velleman, David. 1989. *Practical Reflection*. Princeton N.J.: Princeton University Press.

———. 1992. "What Happens When Someone Acts?" *Mind* 101 (403). New Series (July 1): 461–481.

———. 2000. *The Possibility of Practical Reason*. Oxford;New York: Clarendon Press; Oxford University Press.

Wegner, Daniel M. 2003. *The Illusion of Conscious Will*. Cambridge, Mass.: MIT Press.

Weintraub, Ruth. 1987. "Unconscious Mental States." *The Philosophical Quarterly* 37 (149) (October 1): 423–432.

Westphal, K. R. 1991. "Hegel's Critique of Kant's Moral World View." *Philosophical Topics* 19 (2): 133–176.

Willaschek, Marcus. 1991. "Die Tat Der Vernunft. Zur Bedeutung Der Kantischen These Vom 'Factum Der Vernunft'." In *Akten Des VII. Internationalen Kant-Kongresses Mainz 1990*, Bd. II.1:456–66. Bonn.

Williams, Garrath. 2009. "Kant's Account of Reason." Ed. Edward N. Zalta. *The Stanford Encyclopedia of Philosophy*.
http://plato.stanford.edu/archives/sum2009/entries/kant-reason/.

Wilson, George, and Samuel Shpall. 2012. "Action." In *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta. Summer 2012.
http://plato.stanford.edu/archives/sum2012/entries/action/.

Wittgenstein, Ludwig. 2003. *Philosophische Untersuchungen*. 1. Aufl. Frankfurt am Main: Suhrkamp.

Woodward, James. 2008. "Causation and Manipulability." In *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta. Winter 2008.
http://plato.stanford.edu/archives/win2008/entries/causation-mani/.

Wright, Larry. 1972. "Explanation and Teleology." *Philosophy of Science* 39 (2) (June 1): 204–218.

———. 1973. "Functions." *The Philosophical Review* 82 (2) (April 1): 139–168.

Zhu, J., and P. Thagard. 2002. "Emotion and Action." *Philosophical Psychology* 15 (1): 19–36.

Samenvatting

0 - Algemene Introductie

Een deel van ons gedrag is reflexmatig, of doen we uit gewoonte. Dat we met onze ogen knipperen wanneer de zon er recht in schijnt hangt niet af van een bewuste keuze, en dat we remmen als een voetganger voor onze auto opduikt is ook niet iets waar een bewuste keuze aan vooraf gaat. Een ander deel van ons gedrag is een stuk minder reflexmatig, maar soms net zo onbewust. Een voorbeeld is het drinken van een kopje koffie tijdens avond werkzaamheden. Zoiets doe je om alert te blijven. Natuurlijk proef je de koffie die je drinkt, maar je bent je er niet altijd *expliciet* bewust van dat je koffie drinkt - laat staan dat jij je er explicit bewust van bent *dat je koffie drinkt om alert te blijven*. Natuurlijk is het mogelijk dat jij je er expliciet bewust van bent dat je koffie drinkt om alert te blijven. Een dergelijke vorm van bewustzijn dat je iets doet met een reden noem ik 'actoraal zelfbewustzijn'. Van een deel van ons gedrag weten we, terwijl we het doen, *wat* we doen en *waarom*. Normaal gesproken zeggen we van zulk gedrag dat je het bewust, met opzet en intentioneel doet.

In mijn dissertatie onderzoek ik 'de ongebruikelijke stelling' dat actoraal zelfbewustzijn onafhankelijk is van het *vermogen* bewust te handelen met een reden. Volgens de *ongebruikelijke stelling* zou het, bijvoorbeeld, mogelijk kunnen zijn dat iemand zich bewust is dat hij koffie drinkt om alert te blijven *onafhankelijk* van een vermogen deze of een andere rationele handeling bewust te verrichten. Ik *ontwikkel* de ongebruikelijke stelling met als doel deze voor verder onderzoek op de filosofische agenda te plaatsen, d.i. zonder enige pretentie de stelling te onderschrijven of te bewijzen. De ongebruikelijke stelling gaat over actoraal zelfbewustzijn als zodanig, en laat open of wij, mensen, in staat zijn bewust te handelen met een reden. De stelling propageert uitsluitend dat een dergelijk vermogen, mochten wij het hebben, niet intrinsiek verbonden is met actoraal zelfbewustzijn.

De ongebruikelijke stelling, hoewel tegen-intuïtief, is niet zo vreemd als deze op het eerste gezicht lijkt. Denk bijvoorbeeld eens aan een kat die naar zijn voerbak loopt. Het is vrij gebruikelijk te zeggen dat de kat dit doet omdat dit hem bij zijn eten brengt. Tegelijkertijd is het net zo gebruikelijk te ontkennen dat de kat zich er expliciet bewust van is dat hij naar zijn voerbak loopt om te eten. Denk vervolgens eens aan iemand die zich er bewust van is dat hij rookt om te ontspannen, maar tegelijkertijd op grond van dit bewustzijn niet in staat lijkt te stoppen met roken. Kunnen we dit concept van het soort handelingsvermogen dat we geneigd zijn aan de kat toe te schrijven combineren met een concept van het type bewustzijn dat de roker heeft, om op deze manier de ongebruikelijke stelling te ontwikkelen?

1 – Louter Actoraal Zelfbewustzijn?

Wat moeten we ons voorstellen bij actoraal zelfbewustzijn dat onafhankelijk is van het vermogen bewust te handelen met een reden? Neem bijvoorbeeld de koffiedrinker die actoraal zelfbewustzijn heeft. Hij kan zich op twee verschillenden manieren van zichzelf en zijn handeling bewust zijn. Hij kan denken 'ik drink koffie om alert te blijven, ben me daar bewust van, en blijf koffie drinken in het licht van dit bewustzijn'. Bij wijze van alternatief kan hij echter ook denken 'ik drink koffie om alert te blijven en ben me daar bewust van'. Dit alternatief wordt geïmpliceerd door de eerdere denkwijze, maar zou de eerdere denkwijze enkel impliceren als de actor er van uitgaat dat hij in staat is op grond van zijn bewustzijn te handelen. De actor hoeft dit, op het eerste gezicht, niet te veronderstellen. Als hij dit doet en niet hoeft te doen, dan zou hij 'louter' actoraal zelfbewustzijn hebben.

Om de stelling dat *louter* actoraal zelfbewustzijn mogelijk is te ontwikkelen moet de 'twee capaciteiten stelling' ontwikkeld worden. Deze stelling is dat (i) het vermogen te handelen met een reden onafhankelijk is van het vermogen bewust te handelen met een reden; en (ii) dat het vermogen tot actoraal zelfbewustzijn onafhankelijk is van het vermogen bewust te handelen met een reden. Om deze twee deelstellingen te ontwikkelen, interpreteer ik ze tegen de achtergrond van een onderscheid dat David Velleman maakt in zijn model van de praktische rede. Velleman propageert een onderscheid tussen het *praktische* vermogen van een actor tot onbewuste doelgerichte handelingen en het *epistemische* vermogen van de actor zich een zelfbegrip te vormen. Deze twee vermogens van een actor maken volgens Velleman de praktische rede mogelijk, wanneer de actor daarnaast ook de *dispositie* heeft alleen zo te handelen dat hij kan begrijpen wat hij doet terwijl hij het doet. Daarom concludeer ik dat deze twee vermogens, in Velleman's model, onafhankelijk zijn van de praktische rede. Om op basis van Velleman's onderscheid de twee-capaciteiten stelling te ontwikkelen is er een model nodig waarin enerzijds het vermogen te handelen met een reden, en anderzijds het vermogen van een actor expliciet te maken dat hij handelt met een reden, beide onafhankelijk zijn van het vermogen bewust te handelen met een reden.

2 – Op zoek naar een Royaal Concept van Rationeel Actorschap

Het praktische vermogen te handelen met een reden, wordt paradigmatisch begrepen als een bewust handelingsvermogen. Toch gaan er in de literatuur stemmen op dat een dergelijk bewustzijn niet zo reflectief en conceptueel hoeft te zijn, zoals het verondersteld wordt te zijn in het geval van mensen. Met andere woorden, er is een pleidooi voor een 'royaal' concept van rationeel actorschap, d.i. een concept waarbij rationale handelingscapaciteiten weliswaar mentale, maar pre-reflectieve en pre-conceptuele capaciteiten zijn. Als zodanig lijkt dit royale concept een stap te zijn in de richting van de twee capaciteiten stelling. Toch is dit royale concept in de huidige context problematisch, omdat het sterk suggereert dat onbewust rationeel actorschap alleen mogelijk is in het geval van een actor wiens bewustzijn pre-conceptueel en pre-reflectief is. Echter, om de twee capaciteiten stelling te ontwikkelen moeten we juist onbewust rationeel actorschap kunnen toeschrijven aan een rationele actor die een actoraal, d.i., expliciet conceptueel, zelfbewustzijn heeft. Daarom hebben we dus een ander royaal concept van rationeel actorschap nodig, d.i. een concept dat open laat of het vermogen te handelen met een reden mentaal is. Een noodzakelijke eerste stap, in de richting van een dergelijk royaal concept, is om redenen niet langer te begrijpen als

mentale toestanden (zoals verlangens, intenties, besluiten, keuzes etc.) van de actor. Bijvoorbeeld dat de reden waarom je koffie drinkt niet je intentie is wakker te blijven, maar het feit dat je wakker blijft als je koffie drinkt. Met Elisabeth Anscombe kunnen we dergelijke handelingen begrijpen als handelingen die weliswaar een waarom-vraag voor redenen mogelijk maken, maar we moeten de mogelijkheid onderzoeken dat deze redenen geen mentale toestanden hoeven te zijn.

3 – Een Onbewuste Rationele Handelingscapaciteit?

Charles Taylor heeft beargumenteerd dat een handeling, in tegenstelling tot andersoortige gebeurtenissen, gemarkeerd wordt door het feit dat deze een (bepaald soort) *teleologische* verklaring nodig heeft. Een teleologische verklaring legt een gebeurtenis *x* uit, als iets dat gebeurt omdat *x noodzakelijk* is voor het tot stand komen van *s.* De noodzaak van *x* voor het tot stand komen van *s,* is de *oorzaak* van *x*. Deze verklaring is teleologisch, omdat je kunt zeggen dat *s* het *doel* van *x* is. Bijvoorbeeld het hart van een organisme pompt bloed, omdat het pompen van bloed nodig is voor de zuurstoftoevoer door zijn hele lichaam. Je zou kunnen zeggen dat zuurstoftoevoer het doel is van het pompen van het bloed. Volgens Taylor moet een handeling uitgelegd worden als iets dat een actor doet vanwege de doelen die hij heeft. Je drinkt koffie omdat je als doel hebt wakker te blijven. Hoewel Taylor's model van teleologische verklaring in grote lijnen bruikbaar zal blijken om uit te leggen dat redenen geen mentale toestanden hoeven te zijn, is zijn model van handelingsverklaring daarvoor onbruikbaar. Dit is het geval, omdat Taylor er van uitgaat dat een handeling doelmatig is, en deze doelmatigheid uitlegt als iets dat komt door een actor die op grond van zijn verlangens en intenties vorm geeft aan zijn gedrag. Met andere woorden, Taylor beroept zich op het idee dat handelingen voortkomen uit het bewustzijn van een actor.

Taylor's claim dat handelingen een bepaald type teleologische verklaring, enerzijds, en zijn analyses van teleologische verklaringen en verklaringen van handelingen, anderzijds, laten ruimte om de stelling te ontwikkelen dat handelingen niet inherent verbonden zijn met het bewustzijn van een actor. Ik ontwikkel de stelling dat handelingen gedefinieerd worden door het type verklaring dat ze nodig hebben: een handeling x_1 vindt plaats, omdat een actor *motivationeel responsief* is ten opzichte van het *feit* dat x_1 een factor is voor *s.* Koffie drinken om alert te blijven, moet uitgelegd worden als iets dat een actor doet omdat hij motivationeel responsief is ten opzichte van het feit dat het drinken van deze koffie een factor is voor hem om alert te blijven.

Dergelijke *feiten* zijn fysieke, niet mentale, *redenen*. Verder is het nog maar de vraag of dergelijke motivationele responsiviteit bewustzijn veronderstelt. Ik beargumenteer dat als x_1 is geïnitieerd - en onafhankelijk van hoe x_1 geïnitieerd werd - er al sprake kan zijn van een dergelijke motivationele responsiviteit. Dit is het geval op het moment dat de toestand van het motivationele systeem die de beweging x_1 in gang houdt dan wel stopt, daarvan afhangt of x_1 feitelijk een factor is voor s. Een dergelijke motivationele responsiviteit is mogelijk, onafhankelijk van mentale capaciteiten.

4 – Motivationele Responsiviteit of Praktisch Oordeelsvermogen

In dit hoofdstuk pareer ik de gebruikelijk filosofische stelling dat handelen met een reden afhangt van een vermogen tot praktische oordeelsvorming. Tegen de achtergrond van deze stelling worden de redenen waarvoor een actor x_1 doet impliciet begrepen als de gronden waaruit hij concludeert dat het goed is x_1 te doen. Volgens deze stelling zou iemand die koffie drinkt om alert te blijven, dit doen op grond van zijn conclusie dat het goed is koffie te drinken omdat hij alert wil blijven. Ik benadruk dat een uitleg van x_1 in termen van motivationele responsiviteit ten opzichte van het feit dat x_1 een factor is voor s, begrepen kan worden zonder een dergelijk reflectief praktisch vermogen toe te schrijven aan een actor. In het bijzonder, weerleg ik dat een dergelijke uitleg impliciet veronderstelt dat de actor x_1 doet omdat hij s beoogt en in het licht daarvan en van het feit dat x_1 een factor is voor s, concludeert dat x_1 goed is. Tegen de achtergrond van het in hoofdstuk 3 geschetste model van rationeel actorschap, benadruk ik dat redenen van de vorm 'x is een factor voor s' feiten zijn, en onafhankelijk van normen. Een actor voor wie 'x een factor voor s' is, leeft in een wereld waarin het een feit is dat x een factor voor s is. In deze feitelijke omstandigheden heeft de actor een reden x te doen omdat x een factor voor s is. Motivationele responsiviteit ten opzichte van een dergelijk feit is enkel en alleen datgene wat toegeschreven moet worden aan een actor om zijn gedrag uit te leggen in termen van een dergelijk feit. Het hoeft niet geduid te worden als een vermogen tot praktische oordeelsvorming.

Willen we het royale concept van rationeel actorschap weerleggen op grond van de claim dat rationele actoren in staat zijn praktische oordelen te vellen, dan moeten we hard kunnen maken dat rationele actoren een bewust vermogen hebben praktische oordelen te vellen. De stelling dat rationele actoren een dergelijk bewust vermogen hebben, moet dan onderbouwd worden vanuit een meer algemene analyse van het bewustzijn van rationele actoren. In de volgende twee hoofdstukken ontwikkel ik de

stelling dat het vermogen van een actor om expliciet te maken dat hij handelt met een reden een rudimentair talige capaciteit is, die onafhankelijk is van een vermogen tot bewust rationeel handelen.

5 – Expliciet Maken dat Je Handelt met een Reden

Om louter actoriaal zelfbewustzijn van een actor uit te leggen in termen van zijn vermogen om expliciet te maken dat hij handelt met een reden, moet ik minstens uitleggen dat dit vermogen zelf geen bewust rationeel handelingsvermogen is *en* er ook niet van afhangt. In deze context lees ik Ludwig Wittgensteins 'private language argument' als een argument voor de stelling dat zelfbewustzijn afhangt van inherent sociale talige vermogens, maar ontwikkel de stelling dat deze talige vermogens zelf onbewust zouden kunnen zijn. Om dit uit te leggen val ik terug op George Herbert Mead's model van taal. Mead analyseert het vermogen om taal te gebruiken, in zekere zin, in termen van het vermogen van een lid van een sociale groep om een gebaar te maken dat in andere leden een ander gebaar oproept. Een dergelijk gebaar heeft daarmee een betekenis, maar is enkel een symbool als een actor deze maakt om daarmee een bepaalde reactie op te wekken in de groep. Ik leg uit dat het vermogen symbolische gebaren te maken, hoewel het zelfbewustzijn mogelijk kan maken, zelf geen bewust vermogen hoeft te zijn. In termen van Mead's analyse, zou een actor het vermogen hebben om expliciet te maken dat hij handelt met een reden, als hij in staat is een gebaar te maken om daarmee in de groep de reactie op te wekken er mee in te stemmen dat hij handelt met die reden.

Daarnaast ontwikkel ik de stelling dat het vermogen van een actor om expliciet te maken dat hij handelt met een reden, zelf niet afhangt van een vermogen bewust te handelen met een reden. Als een actor expliciet maakt dat hij koffie drinkt om alert te blijven is dat, net als het drinken van koffie om alert te blijven, op zichzelf een handeling. Deze handeling is impliciet een reactie op *weer* een andere handeling, namelijk het stellen van de vraag 'waarom drink ik koffie?' Hoewel ik denk dat al deze handelingen rationeel (kunnen) zijn, beargumenteer ik dat ze net zo min bewust hoeven te zijn als het drinken van koffie om alert te blijven.

6 – Louter Actoraal Zelfbewustzijn

Traditioneel wordt zelfbewustzijn begrepen op een manier die problematisch is met het oog op de twee-capaciteiten stelling. Ofwel er wordt verondersteld dat het

samenvalt met en beperkt is tot het bewustzijn van mentale toestanden. Ofwel zelfbewustzijn wordt begrepen in termen van een inherent bewuste activiteit, die intrinsiek in staat is *en* verantwoordelijk is voor een coherente constitutie van mentale toestanden. Als we actoraal zelfbewustzijn op een van deze manieren willen begrijpen, dan zouden we ofwel moeten veronderstellen dat (bewustzijn van) motivationele responsiviteit ten opzichte van redenen (een bewustzijn van) een mentale toestand is. In dit geval zouden we afscheid moeten nemen van het eerder ontwikkelde royale concept van rationeel actorschap. Ofwel we zouden moeten stellen dat actoraal zelfbewustzijn het resultaat is van een bewuste, rationele activiteit. In dit geval, zou actoraal zelfbewustzijn altijd afhangen van een vermogen bewust te handelen met een reden. Om deze twee problemen te omzeilen grijp ik terug op Tugenhat's model van zelfbewustzijn als *observatie onafhankelijke zelfkennis*. Volgens Tugendhat is dergelijk zelfbewustzijn alleen mogelijk voor een actor wiens mentale toestanden waarneembaar zijn in de taalgemeenschap waartoe de actor behoort. Een dergelijke actor kan het vermogen geleerd worden zichzelf door het gebruik van 'ik' aan te wijzen als een identificeerbaar lid van deze taalgemeenschap; en in combinatie daarmee bepaalde talige expressies te gebruiken die binnen deze taalgemeenschap mentale toestanden toeschrijven aan het met 'ik' aangewezen identificeerbare lid. Omdat deze mentale toestanden natuurlijke expressies hebben zijn ze waarneembaar, en op grond daarvan kan een actor geleerd worden expressies zoals 'ik heb pijn' te gebruiken *dan en slechts dan als*, binnen zijn taalgemeenschap, de natuurlijke expressie van zijn pijn waarneembaar is.

Ik claim vervolgens dat Tugendhat's argument ook opgaat in het geval van niet-mentale toestanden. Zolang een actor op bovenstaande wijze geleerd kan worden – onafhankelijk van zelfobservatie – talig expressie te geven aan zijn toestand die waarneembaar is binnen deze taalgemeenschap, kan hij zonder zelfobservatie kennis hebben van deze toestand *onafhankelijk of dit een mentale toestand is*. Motivationele responsiviteit ten opzicht van redenen kan een dergelijke toestand zijn. Een dergelijk argument stelt mij in staat de tweede component van de twee-capaciteiten stelling te ontwikkelen, zonder daarbij mijn uitleg van de eerste component in gevaar te brengen.

7 – Algemene Conclusie

Als de ongebruikelijke stelling dat actoraal zelfbewustzijn onafhankelijk is van een vermogen bewust te handelen met een reden waar is, dan zou dit verschillende

implicaties hebben. Het zou vooral betekenen dat we nog eens goed moeten discussiëren over de vraag of, en zo ja waarom en in welke mate, wij er vanuit moeten gaan dat we in staat zijn bewust te handelen voor een reden. Deze discussie is belangrijk in tenminste drie opzichten. Als *eerste* biedt het een nieuw perspectief op de vraag of en de mate waarin wij onze handelingen moeten conceptualiseren als dingen waar we op grond van onze veronderstelde vrijheid verantwoordelijkheid voor zijn. Als *tweede* roept het vragen op over de gronden waarop en het feit dat we aan onszelf een morele status toekennen, waarmee we onze belangen verheffen boven die van actoren zonder actoraal zelfbewustzijn en boven de hoedanigheid van andere natuurlijke entiteiten. Deze conventionele en geïnstitutionaliseerde zelfverheffing is, naar mijn mening, de meest belangrijke oorzaak voor de huidige situatie waarin we vrede lijken te hebben met het feit dat onze manier van leven afhankelijk is van de uitbuiting van bepaalde bevolkingsgroepen en dieren, maar ook nauw verbonden met de uitputting en destructie van de natuurlijke leefomgeving van een huidige en toekomstige generatie van mensen en overige levensvormen. Als *derde* biedt deze discussie een nieuwe basis voor de vraag of we er in het westen juist aan doen mensen, op grond van hun veronderstelde vermogen zelf afwegingen en keuzes te kunnen maken, proberen op te voeden en te corrigeren als individuen die verantwoordelijk zijn voor hun eigen gedrag en positie in de samenleving.

In mijn dissertatie heb ik geen poging ondernomen te bewijzen dat de ongebruikelijke stelling waar is, alleen dat deze stelling ontwikkeld kan worden in het licht van verschillende aspecten van verschillende filosofische theorieën. Gezien wat er van af zou hangen mocht deze stelling waar zijn, is het mijn suggestie – in het licht van het door mij verrichte voorwerk - deze ongebruikelijke stelling verder te onderzoeken niet alleen op coherentie maar ook op waarheid.

Curriculum Vitae

Gerhard Bos was born in Kampen on October 4, 1983. After one year of studying Information Technology at University of Twente, he completed his Bachelor's and Research Master's in Philosophy at Utrecht University. In 2006 he finished his BA with a thesis in meta-ethics on coherentist approaches to moral justification, which was awarded as the best bachelor's thesis written at the Faculty of Philosophy. His master's thesis on Kant's philosophy of freedom, which we wrote under supervision of Prof. dr. Deryck Beyleveld in 2008, was awarded with the Best Thesis Award of Utrecht University. Since 2004 he was involved teaching logic, philosophy of language and ethics. From 2008-12 he worked as a PhD-student at the Ethics Institute of Utrecht University, doing research and teaching ethics to philosophy and non-philosophy bachelor's and master's students. His dissertation was supervised by Prof. dr. Marcus Düwell, Dr. Thomas Müller and Dr. Micha Werner. Since 2010 he is member of the European Platform for the Life Sciences, Mind Sciences and the Humanities, a platform of the Volkswagen Foundation. In winter 2010-11 he went to the Westfälische Wilhelms-Universität Münster, to attend to seminars on Marx and Hegel. Since November 2012 he works as a Postdoc at Utrecht University doing research on conceptual problems with ascribing rights to future generations.

Quaestiones Infinitae

PUBLICATIONS OF THE DEPARTMENT OF PHILOSOPHY

1998.

VOLUME 27. M.F. VERWEIJ, *Preventive Medicine Between Obligation and Aspiration* (dissertation), 1998.

VOLUME 28. J.A. BERGSTRA, S.F.M. VAN VLIJMEN, *Theoretische Software-Engineering: kenmerken, faseringen en classificaties*, 1998.

VOLUME 29. A.G. WOUTERS, *Explanation Without A Cause* (dissertation), 1999.

VOLUME 30. M.M.S.K. SIE, *Responsibility, Blameworthy Action & Normative Disagreements* (dissertation), 1999.

VOLUME 31. M.S.P.R. VAN ATTEN, *Phenomenology of choice sequences* (dissertation), 1999.

VOLUME 32. V.N. STEBLETSOVA, *Algebras, Relations and Geometries (an equational perspective)* (dissertation), 2000.

VOLUME 33. A. VISSER, *Het Tekst Continuüm* (inaugural lecture), 2000.

VOLUME 34. H. ISHIGURO, *Can we speak about what cannot be said?* (public lecture), 2000.

VOLUME 35. W. HAAS, *Haltlosigkeit; Zwischen Sprache und Erfahrung* (dissertation), 2001.

VOLUME 36. R. POLI, *ALWIS: Ontology for knowledge engineers* (dissertation), 2001.

VOLUME 37. J. MANSFELD, *Platonische Briefschrijverij* (valedictory lecture), 2001.

VOLUME 37A. E.J. BOS, *The Correspondence between Descartes and Henricus Regius* (dissertation), 2002.

VOLUME 38. M. VAN OTEGEM, *A Bibliography of the Works of Descartes (1637-1704)* (dissertation), 2002.

VOLUME 39. B.E.K.J. GOOSSENS, *Edmund Husserl: Einleitung in die Philosophie: Vorlesungen 1922/23* (dissertation), 2003.

VOLUME 40.   H.J.M. BROEKHUIJSE, *Het einde van de sociaaldemocratie* (dissertation), 2002.

VOLUME 41.   P. RAVALLI, *Husserls Phänomenologie der Intersubjektivität in den Göttinger Jahren: Eine kritisch-historische Darstellung* (dissertation), 2003.

VOLUME 42.   B. ALMOND, *The Midas Touch: Ethics, Science and our Human Future* (inaugural lecture), 2003.

VOLUME 43.   M. DÜWELL, *Morele kennis: over de mogelijkheden van toegepaste ethiek* (inaugural lecture), 2003.

VOLUME 44.   R.D.A. HENDRIKS, *Metamathematics in Coq* (dissertation), 2003.

VOLUME 45.   TH. VERBEEK, E.J. BOS, J.M.M. VAN DE VEN, *The Correspondence of René Descartes: 1643*, 2003.

VOLUME 46.   J.J.C. KUIPER, *Ideas and Explorations: Brouwer's Road to Intuitionism* (dissertation), 2004.

VOLUME 47.   C.M. BEKKER, *Rechtvaardigheid, Onpartijdigheid, Gender en Sociale Diversiteit; Feministische filosofen over recht doen aan vrouwen en hun onderlinge verschillen* (dissertation), 2004.

VOLUME 48.   A.A. LONG, *Epictetus on understanding and managing emotions* (public lecture), 2004.

VOLUME 49.   J.J. JOOSTEN, *Interpretability formalized* (dissertation), 2004.

VOLUME 50.   J.G. SIJMONS, *Phänomenologie und Idealismus: Analyse der Struktur und Methode der Philosophie Rudolf Steiners* (dissertation), 2005.

VOLUME 51.   J.H. HOOGSTAD, *Time tracks* (dissertation), 2005.

VOLUME 52.   M.A. VAN DEN HOVEN, *A Claim for Reasonable Morality* (dissertation), 2006.

VOLUME 53.   C. VERMEULEN, *René Descartes, Specimina philosophiae: Introduction and*

Critical Edition (dissertation), 2007.

VOLUME 54.   R.G. MILLIKAN, *Learning Language without having a theory of mind* (inaugural lecture), 2007.

VOLUME 55.   R.J.G. CLAASSEN, *The Market's Place in the Provision of Goods* (dissertation), 2008.

VOLUME 56.   H.J.S. BRUGGINK, *Equivalence of Reductions in Higher-Order Rewriting* (dissertation), 2008.

VOLUME 57.   A. KALIS, *Failures of agency* (dissertation), 2009.

VOLUME 58.   S. GRAUMANN, *Assistierte Freiheit* (dissertation), 2009.

VOLUME 59.   M. AALDERINK, *Philosophy, Scientific Knowledge, and Concept Formation in Geulincx and Descartes* (dissertation), 2010.

VOLUME 60.   I.M. CONRADIE, *Seneca in his cultural and literary context: Selected moral letters on the body* (dissertation), 2010.

VOLUME 61.   C. VAN SIJL, *Stoic Philosophy and the Exegesis of Myth* (dissertation), 2010.

VOLUME 62.   J.M.I.M. LEO, *The Logical Structure of Relations* (dissertation), 2010.

VOLUME 63   M.S.A. VAN HOUTE, *Seneca's theology in its philosophical context* (dissertation), 2010.

VOLUME 64   F.A. BAKKER, *Three Studies in Epicurean Cosmology* (dissertation), 2010.

VOLUME 65   T. FOSSEN, *Political legitimacy and the pragmatic turn* (dissertation), 2011.

VOLUME 66   T. VISAK, *Killing happy animals. Explorations in utilitarian ethics.* (dissertation), 2011.

VOLUME 67   A. JOOSSE, *Why we need others: Platonic and Stoic models of friendship and self-understanding* (dissertation), 2011.

VOLUME 68   N. M. NIJSINGH, *Expanding newborn screening programmes and strengthening informed consent* (dissertation), 2012.

VOLUME 69   R. PEELS, *Believing Responsibly: Intellectual Obligations and Doxastic Excuses* (dissertation), 2012.

VOLUME 70   G.H. BOS, *Agential Self-consciousness, beyond conscious agency* (dissertation), 2013.