UNIVERSITÀ DEGLI STUDI
DI MILANO

Facoltà di Lettere e Filosofia
Scuola di Dottorato in Humanae Litterae
Dottorato di Ricerca in Filosofia

Motor intentions:
connecting intentions with actions

Tesi di dottorato di:
Chiara Brozzo
Matr. n. R08600

Tutor: Prof. A. Zucchi
Prof. S. Butterfill
Prof. C. Sinigaglia
Coordinatore del dottorato: Prof. P. Spinicci

XXV Ciclo

« L'avenir n'est jamais que du présent à mettre en ordre. Tu n'as pas à le prévoir, mais à le permettre. »

Antoine de Saint-Exupéry, *Citadelle* (1948)

Contents

Acknowledgements **4**

Introduction **5**

§1. What (if anything) do mirror neurons tell us about intentions? **17**

Introduction **17**
1. Do mirror neurons tell us about mere movements? **19**
2. Outcomes: more than mere movements **24**
3. The intention interpretation and its attractiveness **31**
3.1 How do we shift from outcomes to intentions? What changes? 31
3.2 What rests on interpreting mirror neuron activation in terms of intentions? 35
3.2.1 What is mindreading? How does it connect with action understanding? 36
3.2.2 Simulation Theory and simulation-based action understanding 38
3.2.3 What is the relation between simulation and mindreading? 41
3.2.4 What is the role of mirroring in action understanding? Causal vs. constitutive role 43
3.2.5 How the intention interpretation could help identifying the role of action mirroring in action understanding 49
3.3 Some provisional conclusions 52
4. Which intentions? How many intentions? **54**
5. Mirror neuron activation: high or low? **56**
Conclusion: motor outcomes—not so high, not so low **66**

§2. What is an intention? **69**

Introduction **69**
1. What is an intentional action? A methodological claim on relating intentional action to responsibility attribution **70**
2. First attempt: reducing intention to intentional action **74**
3. Second attempt: intentions are just desires **80**
4. Connecting intentions with actions via the Simple View and the Single Phenomenon View **83**
4.1 Why intentions are not just desires: Bratman on intentions as plans 83
4.2 Why the Simple View doesn't work (I): the video-games case 89
4.3 Why the Simple View doesn't work (II): webs of norms 93
6. "What good is a will?" **96**

§3. How are intentions and motor representations related? **103**

Introduction **103**
1. What are motor representations and why think they could be intentions? **106**
2. Are motor representations intentions? **109**

3. What's the relation between the content of intentions and the content of motor representations? 115
3.1 How P-intentions and M-intentions are supposed to come apart 118
3.2 How far can awareness of action performance go? 121
3.3 A proposal: one kind of intention, many possible content/consciousness specifications 123
3.4 Intention and fine-tuning of action: two of a kind? 127
Conclusion 130

§4. What is the content of a motor intention? 132

Introduction 132
1. Are motor intentions really intentions? 133
2. Action hierarchies and variable content specification 135
3. Is the content of motor intentions conceptual? 138
4. What is the content of motor intentions? 143
Conclusion 145

§5. Prior intentions vs. intentions in action: exploding the myth 147

Introduction 147
1. The alleged differences between prior intentions and intentions in action: temporal characteristics and content 147
2. Annihilating the differences, exploding the myth 152
2.1 Temporal differences 152
2.2 Differences in content 154
Conclusion 156

§6. What are intentions-with-which? The hierarchical structure underlying intentions 158

Introduction 158
1. Three notions of *intention-with-which* 158
2. Hierarchically structured actions 161
2.1 "A fine-grained procedure for individuating acts" 161
2.2 What is level-generation? Is there a more useful notion for structuring actions? 163
3. Are actions fulfilling motor intentions the same as basic actions? 170
4. Structuring intentions on the basis of actions 179
5. Re-interpreting experimental results in the light of a hierarchical structure for intentions 182
5.1 Re-interpreting results by resorting to the vertical dimension only 182
5.2 Re-interpreting results by resorting to the horizontal dimension only 189
5.3 Re-interpreting results by combining the vertical dimension with the horizontal dimension 192

Conclusion: motor intentions—not so high, not so low 196

References 200

Acknowledgements

Heartfelt thanks to: Margherita Arcangeli, Brian Ball, Luca Barlassina, Tim Bayne, Mark Baynes, Claudio Bergamini, Tommaso Bruni, Stephen Butterfill, Flavia Cardini, Marcello Costantini, Michele Di Francesco, Vittorio Gallese, Pierre Jacob, Angelica Kaufmann, Hemdat Lerman, Matthew Longo, Matteo Mameli, Kirk Michaelian, the Neurophilosophy research group, Krisztina Orban, Sara Parmigiani, Marco Penna, Johannes Roessler, Marco Santambrogio, Corrado Sinigaglia, Barry C. Smith, Joulia Smortchkova, Paolo Spinicci, Sebo Uithol, Maurizio Vitagliani, the PhD students at the Warwick Philosophy Department (in particular Juan Camilo Espejo-Serna, Pete Fossey, Vivan Joseph, Roberta Locatelli, Nicola Spinelli), Hong Yu Wong, Silvano Zipoli Caiani, Sandro Zucchi, my mother and father, all my other family members, and my friends.

Introduction

It's a platitude that we're generally interested in understanding what other people are doing. Instances of sheer curiosity aside, if we lacked such an understanding we wouldn't be able to interact with other people in the right way. If you're holding an object up in the air, I must be able to tell whether you want to hand it to me, in which case I'll come closer and extend my hand to receive it, or whether, instead, you want to throw it to me and hit me with it, in which case I'd better run away and duck. Now, what does it take to understand what other people are doing? One could think that, in order to understand it, all we have to do is observe other's actions. But is this true? Do other people's actions, conceived as their outward bodily movements, suffice to tell what these people are up to? At least in principle, action observation might not suffice. Let us reflect on the following example:

> Consider Dr Jekyll and Mr Hyde. The former is a renowned surgeon who performs appendectomies on his anesthetized patients. The latter is a dangerous sadist who performs *exactly the same hand movements* on his non-anesthetized victims. As it turns out, Mr Hyde is Dr Jekyll.
> (Jacob & Jeannerod, 2005, p. 23, my emphasis)

In order to understand what the surgeon is really doing, according to Jacob and Jeannerod, observing his bodily movements won't suffice, as they're the same both in the case in which the surgeon intends to harm the patient or when he intends to help him. Thus, if a person were to be brought in a law court, and a witness were to say that she saw this person performing certain bodily movements

with a scalpel, independently of any information as to whether this person intended to harm the patient, the witness's report wouldn't help the judge to assess whether the accused person is to be blamed or praised. In a nutshell, if Jacob and Jeannerod are right, observing others' actions might not be enough to understand what they are doing. What else do we need? Here is a plausible answer: we have to understand their intentions. For example, once we're told that the surgeon intended to harm his patient, we are in a position to consider him blameworthy.

But, in order to understand what a person is up to, is it enough to be told that, besides performing a certain action, this person also had a certain intention? Consider the following example due to Davidson (1973). A climber who wants to get rid of the weight and danger of holding another man on a rope might be so unnerved by these reflections as to end up loosening his hold. So the man intends to loosen his hold, and does perform an action of loosening his hold. But loosening his hold, in this case, is not brought about in order to fulfil the corresponding intention. Rather, it just happens to him.

This example brings us the heart of my thesis: in order to understand what a person is doing, we have to understand what action she's performing. Not only that, but we also have to understand what intention lies behind her action. Not only that, but we must also be assured that an action is harnessed to the corresponding intention in the appropriate way: the climber in Davidson's example ends up executing the bodily movements that correspond to his intention, but, even at first blush, you'll be able to say that he didn't quite act in accordance with his intention—his intention and his action, supposing that loosening one's hold as a consequence of nervousness can be called an action, are not connected in

the appropriate way. But just what *is* the appropriate way? How do we characterize it? Answering these questions amounts to answering the central question of the present work: how are actions connected with intentions?

Searle (1983) has done some influential work towards clarifying the appropriate way in which an intention and an action must be connected in order to be able to say that an action has been *intentionally* performed by the subject, by putting forward a theory that features two kinds of intentions: prior intentions and intentions in action. Prior intentions are characterized as intentions that are formed before the initiation of an action, whereas intentions in action are conceived as the causes of the bodily movements by means of which a certain action is executed. What makes the climber's loosening his hold fail to be an intentional action, by Searle's lights, is that the movements of the climber are not under his guidance in the sense that the climber *doesn't have the relevant intention in action*. So, in Searle's theory an intention in action is what connects an intention to an action, and ensures that an executed action is intentional.

My answer to the central question of the thesis, namely the question as to how intentions and actions are connected, if only to an extent builds on an insight of Searle's theory. The latter is certainly on the right track insofar as it aims to pinpoint what ensures the appropriate connection between an intention and the corresponding action, but it still suffers from a fundamental shortcoming: the notion of intention in action, as it stands, is no more than a theoretical placeholder for what makes it the case that an action is intentional (see Pacherie, ms.). Another shortcoming is the emphasis that Searle's theory puts on a temporal difference between intentions that are formed prior to the execution of an action and intentions whose lifespan roughly coincides with that of the corresponding action.

In this thesis, I'm going to elaborate on the notion of intention in action by trying to make it more than a theoretical placeholder. In order to do this, I'm going to draw on the work of Elisabeth Pacherie, insofar as I take it to constitute the best philosophical attempt to combine the search for a connection between intentions and actions with the relevant results in the cognitive neuroscience of action. In this latter field, I'm going to explore mainly the notions of mirror neurons for action and of motor representation, in order to answer the question as to how intentions and actions are connected by means of the introduction of the notion of *motor intention*.

Mirror neurons for action are neurons in the human brain (first discovered in the macaque monkey brain) that fire both when an individual performs a certain action and when she observes someone else performing the same action (Di Pellegrino et al., 1992; Rizzolatti et al., 1996; Gallese et al., 1996; Rizzolatti & Sinigaglia, 2008; Rizzolatti & Sinigaglia, 2010; Mukamel et al., 2010). I'm going to be interested in mirror neurons for action insofar as they enjoy some sort of double life, in the following sense. On the one hand, their activation seems to correlate with bodily movements. On the other hand, as a series of experiments that I'm going to report show, their activation significantly abstracts away from the details of the movements in correlation with which these neurons fire. I'm going to show that the characteristics of their activation gives us grounds for hypothesizing that mirror neuron firing correlates with a specific kind of intentions. Given the specific kind of outcome that is represented by these intentions, which is bodily in nature, I'm going to term these *motor intentions* (cf. Jeannerod, 1994; Pacherie, 2003, 2006, 2008; Jacob & Jeannerod, 2005), about which I'm going to say something in the following shortly.

A *motor representation* (see Jeannerod, 1994, 2006) consists in the representation of an action outcome that depicts the self in action as a generator of forces, and thus determines the pattern of movements that are going to be executed, thereby driving action execution. Insofar as motor representations play a key role in the production of behaviour, they are obvious candidates for fulfilling the role of intentions in action (something that is suggested by Pacherie, 2000, but see also 2003, 2006, 2008). So I'm going to tackle the question as to whether any motor representations are actually intentions. I argue for the idea that those intentions that I term *motor intentions* can have the same content as some motor representations. Working on this assumption, I develop the idea that the content of motor intentions crucially involves one of a limited set of bodily parts, which has to undergo a specific change—one, for instance, that brings it into a specific relation to a certain object (examples are the intention to grasp, or to tear). Motor intentions have

(a) a bodily content, which is relevantly similar to that of an motor representation, and
(b) their content is related to that of an motor representation either by deferral or by identity, depending on the kind of motor representation in question.

On these grounds, I contend that these intentions are the best candidates for explaining the way in which intentions connect with actions.

My appeal to the notions of motor representations and mirror neurons gives me the chance to make a methodological remark. In what follows, I'm going to rely on the one hand on strict armchair philosophical reflection, by looking at the work

that has been done on intention in analytical philosophy (I'm going to make reference to such authors as Anscombe, Davidson, Goldman, Searle, Bratman, Pacherie and Velleman). On the other hand, I'm also going to rely on recent research in the cognitive neuroscience of action. I conceive of the connection between these two research tools in a twofold way, consisting in the following double influence:

1) (From data in the cognitive neuroscience of action to philosophical reflection.) If two competing theories T_1 and T_2 that make different predictions (which affect a given body of experimental data) can equally be made to stand under the theoretical viewpoint, if (say) T_1 fits a body of data better than T_2, then I'm going to prefer T_1 to T_2.
2) (From philosophical reflection to the cognitive neuroscience of action.) If, on the other hand, a certain philosophical theory (T_1) is found to fare better than another (T_2) under the strictly theoretical viewpoint, and T_1 can be applied to a given body of experimental results while forcing a revision of the interpretation that had previously been given of those results, then I'm going to suggest that this revision should indeed be performed.

It is to be emphasized that even those who are sceptical about any kind of combination between philosophical reflection and experimental practice should still be able to find this piece of work to their liking, insofar as it does rely to a significant extent on strict armchair reflection.

The present thesis is organized as follows.

Chapter 1 introduces the notion of mirror neurons for action, namely neurons that fire both when an individual performs a certain action and when she observes someone else performing the same action. I start out by considering a hypothesis according to which mirror neuron activation correlates with mere movements, and I reject it on the basis of a series of experimental results. I then introduce the idea that mirror neuron activity could correlate with action outcomes, conceived as being more abstract than mere movements insofar as action outcomes enjoy a significant degree of independence with respect to movement details. Building on this idea, I further suggest that mirror neuron activity could correlate with a specific kind of intentions, and I try to locate this proposal within the broader framework of action understanding based on Simulation Theory. I contend that clarifying the correlation between mirror neuron activity and a certain kind of intentions could pave the way for a mirroring-based Simulation Theory of action understanding. I conclude the first chapter by arguing, against Csibra (2007), that the outcomes with which mirror neuron activity correlates are of an intermediate kind, that is bound up with movement details while also significantly abstracting away from fine movement details. This kind of outcomes, which I term *motor outcomes*, are going to be those represented by motor intentions.

Chapter 2 is devoted to beginning to answer the central question in the philosophy of intention, namely that as to what, if anything, gives unity to the three guises in which intentions may appear: intentions for the future, intentional actions, and intentions with which someone acts. This chapter focuses on the link between intentional actions and intentions for the future. I start out with a methodological clarification: as much as it is true that the notion of intentional

action is connected to that of attributions of responsibility, blame or praise, in order to clarify the notion of intentional action we'd better steer clear of these notions. Once that is done, I consider and reject the following hypotheses: intentions may be reduced to intentional actions, and intentions are just desires. By building up on the notion that intentions are mental states that represent action outcomes, I use Bratman's planning theory of intention to show that intentions are distinctive propositional attitudes inserted in a specific web of norms, and I review Bratman's argument to the conclusion that the connection between intentions and intentional actions is given by the so-called Single Phenomenon View: whenever one intentionally *A*s, one has the intention to *B*, and it's not necessarily the case that *A* = *B*. I then raise the challenge proposed by Velleman (2007) to this view: what purpose serves an intention once an intentional action is being performed? In response to this challenge, I introduce the idea that an action may be intentional by virtue of being appropriately connected to the relevant intention.

Chapter 3 introduces the notion of motor representation, and explains why there are reasons, on the one hand, to maintain that motor representations are intentions, and, on the other hand, reasons to think that motor representations are the wrong thing to be identified with intentions. In this chapter I'm going to argue for an intermediate, neutral position according to which the content of some motor representations can be a good guide to the content of some intentions, which, following Jeannerod (1994) and in a different sense with respect to Pacherie's (2003, 2006, 2008) use of the term, I'm going to call *motor intentions*. My stance leaves open the possibility that some motor representations could be intentions, but doesn't commit to this idea. Pacherie proposes the distinction between two

kinds of intention whose function is to initiate, sustain, guide and monitor actions: present-directed (or P-)intentions on the one hand, and motor (or M-)intentions on the other hand, which are supposed to differ, among other things, insofar as the contents of P-intentions are accessible to introspection, while the contents of M-intentions are typically not. I endorse the idea that the content of intentions can encompass all the contents that Pacherie describes as pertaining to P-intentions and at least some of the contents that she describes as belonging to M-intentions, with the exclusion of contents concerning action adjustments. Still, I reject the notion that two kinds of intentions should be distinguished on the basis of accessibility to introspection.

Chapter 4 highlights the existence of a variety of motor intentions, corresponding to different levels of an action hierarchy. While the motor intentions corresponding to the lower levels of the hierarchy bear a significant resemblance to motor images in terms of content, the content of motor intentions higher up in the hierarchy is not so straightforwardly identifiable with that of a motor image. This implies that motor intentions higher up in the hierarchy are less controversially classified as intentions, insofar as they have a propositional format and a conceptual content. As far as motor intentions lower in the hierarchy, the acknowledgement that they are propositional attitudes, and of their having a conceptual content relies on the acceptance of the demonstrative strategy and/or of the notion of executable concepts. What brings together all these different kinds of motor representation is the fact that their content involves reference to sequences of bodily configurations. Now I'm in the position to specify the

similarity in terms of content between motor intentions and motor representations, which was introduced in chapter 3, as follows:

(i) motor intentions at the lower levels of the hierarchy refer to action outcomes by deferring their content to that of a motor representation (see Butterfill & Sinigaglia, 2012) whose content is likely to be a motor image;

(ii) motor intentions at the higher levels of the hierarchy are likely to have the same content as that of corresponding motor representations, which is likely not to consist in a motor image.

Chapter 5 tackles the distinction between prior intentions and intentions in action introduced by Searle (1983) in terms of temporal characteristics and of content. According to the temporal characterization, *prior intentions* are formed prior to the performance of an action, whereas *intentions in action* are present as an intentional action unfolds. I show that the temporal criteria don't suffice to draw the distinction. On the basis of O'Shaughnessy (1991), it can be shown that the prior intention has to last until the action is completed. Hence, after action onset there is an overlap between prior intentions and intentions in action, and one of the two kinds of intention seems to become redundant. Can differences in content preserve the distinction between prior intentions and intentions in action? On the basis of both Searle (1983) and Pacherie (2006, 2008), I try to make a case for the idea that there has to be a difference in content between the two kinds of intention before and after action onset. However, I also provide the example of a case in which the content of the prior intention is as detailed as that of the corresponding intention in action. So, in that case, there is no difference in content between the two kinds of intention. In conclusion, prior intentions and intentions in action are

insufficiently distinguished on the basis of temporal characteristics, and don't always differ in content related features, so that the distinction, as a whole, is inadequately motivated. The take-home message is that there doesn't seem to be any reason to postulate two different kinds of intention, based on a temporal criterion. So, in the framework for intentions proposed in the following chapter, only one kind of intention is going to feature.

Chapter 6 finally puts me in the position to describe intentions in terms of a hierarchical structure of which motor intentions occupy the lowest levels, and non-motor intentions, if any, occupy the levels above. Here I take up again the notion of action hierarchy, so far only hinted at. I elaborate on Goldman's (1970) notion of level-generation among different action descriptions, and I reinterpret it in terms of what I call means-end-generation, which comes at least into two kinds (properly called means-end-generation, and determinate-determinable-generation). Action descriptions can be ordered hierarchically according to the principle of means-end-generation. Once action descriptions have been ordered hierarchically this way, it is possible to obtain a second hierarchical ordering of the intentions corresponding to the action descriptions that feature in the first hierarchy. Thus what I call a vertical dimension for ordering intentions appears, and motor intentions are shown to occupy the lowest level of this hierarchical ordering. This gives me the chance to draw a distinction between the notion of motor intention and that of intention fulfilled by a basic action. I also trace a horizontal dimension for ordering intentions on the basis of temporal relations holding among the corresponding actions. In the light of this hierarchy, it is possible to account for two notions of intention-with-which that are distinguished at the beginning of the

chapter. I conclude by re-interpreting a series of experimental results presented in chapter 1 and some additional ones in the light of both dimensions (vertical and horizontal) of the hierarchical structure for intentions just introduced.

§1. What (if anything) do mirror neurons tell us about intentions?

Introduction

Suppose I'm in the presence of other people who are carrying out some familiar activities—say, someone's bringing food to her mouth, while someone else is moving objects away from table. While I observe these individuals, I understand what they are doing—that is, I understand that the goals of their actions are, respectively, eating and tidying up. What does this understanding consist in? Is it the case that I explicitly reason on the details of the action that I observe to then conclude that it's an instance of, e.g., eating? Do I thereby believe that it is the subject's intention to do it? Or does my action understanding consist in a different process altogether?

Contemporary cognitive neuroscience may have something to contribute to this matter. Consider that there exist neurons in the human brain (first discovered in the macaque monkey brain) that fire both when an individual performs a certain action and when she observes someone else performing the same action, namely mirror neurons[1] (Di Pellegrino et al., 1992; Rizzolatti et al., 1996; Gallese et al., 1996; Rizzolatti & Sinigaglia, 2008; Rizzolatti & Sinigaglia, 2010; Mukamel et al., 2010). It's reasonable to suppose that the firing of mirror neurons in the brain of a

[1] Even though mirror neurons were originally discovered in the domain of action, in recent years a number of studies have established the existence of mirror neurons for emotions and sensations as well (Avenanti et al., 2005; Gallese et al., 2004; Keysers et al., 2004; Singer et al., 2004; Wicker et al., 2003). In the present work, however, I'm going to focus on mirror neurons for actions only—and it's only to those that the provided characterization in terms of action performance and action observation applies.

subject who performs a certain action might serve the purpose of behaviour production. However, the question arises as to what role, if any, the firing of mirror neurons plays in the observer's brain. One might conjecture that this neural activation in each individual's brain enables the individual to represent the intention to perform the action that one of the two is executing (Gallese, 2001), and that mirror neuron activation in the observer is connected to action understanding (Gallese & Goldman, 1998). Some (Rizzolatti et al., 2001, p. 661; Gallese et al., 2004, p. 396) have gone as far as saying that mirror neuron activation in the observer while witnessing another's action *constitutes* understanding of that action. Others (Goldman, 2009, pp. 236-ff.) have more cautiously claimed that mirror neuron activation plays a *causal*, but not *constitutive*, role. Others still have denied that mirror neurons have a "dominant, specialised role in action understanding" (Heyes, 2010). So, what role, if any, does the firing of mirror neurons play in action understanding? So long as answers to these questions cluster around two extremes, consisting of theorists who are either overly enthusiastic or overly pessimistic, no progress will be made in the debate. I contend that a more empirically plausible and theoretically fruitful middle ground for the role of mirror neurons in action understanding should be identified.

In order to do this, some theoretical grounds have to be cleared. In particular, it has to be clarified what exactly mirror neuron activation correlates with[2]. Formulating and defending a specific solution to this issue is what my thesis

[2] The question might be alternatively posed in terms of what mirror neurons represent. I'm going to avoid formulating the question in this way, insofar as there are deep and difficult questions as to whether it can be said that mirror neurons represent anything, and, even if it can be said so, it's unclear what would constitute the relevant unit of representation (a population of neurons vs. a pattern of activation involving a number of neurons, etc.). On questions about representation, see Dretske (1981). For an application of that question to the present issue, see Butterfill & Sinigaglia (2012).

shall be concerned with, in the attempt to answer a more general question as to how the connection between intentions and actions is ensured. Since this is the main question that I mean to address, the following methodological remark has to be made. Notice that the defining property of mirror neurons is that they fire both during action production and during action observation. In the present work, however, I'm mainly going to focus on their role in action production. Still, in order to explore provisional answers to the question as to what mirror neuron activation correlates with, in what follows I'm going to review a series of experimental results concerning mirror neurons that investigate what happens in action observation as much as in action production. Depending on the experimental study under consideration, at times I'll also consider the results obtained as far as action observation is concerned, working on the assumption that, on the basis of the defining property of mirror neurons, these results are going to carry over to what mirror neuron activation correlates with in action production.

1. Do mirror neurons tell us about mere movements?

In a single-cell recording study, Rizzolatti et al. (1988) recorded the activations of neurons in the ventral premotor cortex (area F5) of the macaque monkey brain, while the monkey was performing the following action: reaching for a piece of food, grasping the piece of food and bringing the piece of food to the mouth. The action was repeatedly performed, and it involved different effectors at different times: sometimes grasping was executed with the right hand, sometimes with the left hand, sometimes with mouth. The results were as follows. Some neurons (*Grasping-with-the-hand-neurons*) fired during the act of grasping with the hand, regardless of whether this act was performed with the right or the left hand, and

regardless of the spatial position of the stimulus. Other neurons (*Grasping-with-the-hand-and-with-the-mouth-neurons*) fired during the act of grasping, regardless of whether this act was performed with the hand or with the mouth.

It was then discovered (Di Pellegrino et al., 1992) that a subset of F5 neurons has visual properties too. More specifically, a series of studies (Di Pellegrino et al., 1992; Gallese et al., 1996) established the existence of a class of F5 neurons with the following visual properties: these neurons are not triggered by the observation of objects or mere bodily movements, but only by the observation of actions in which an individual's hand or mouth interacts with an object. Depending on the kind of interaction between effector and object, they were thus classified as follows: *grasping neurons* (responding to the sight of a hand approaching and grasping an object); *placing neurons* (responding to the sight of a hand moving an object on a support); *holding neurons* (responding to the sight of an individual keeping an object in her hand), and so forth. The striking discovery is that these neurons fire in correlation with the same kind of action both during execution and during observation. In a nutshell, their visual properties mirror their motor properties—hence, the name bestowed on these sensorimotor neurons: *mirror neurons*.

So far, I've described mirror neurons recorded in the area F5, but mirror neurons are also present in the inferior parietal lobule (IPL) of the macaque monkey brain (Fogassi et al., 2005). An action-related mirror system is found in the human brain too (Rizzolatti et al., 1996), and is formed mainly by two regions: the inferior section of the precentral gyrus plus the posterior part of the inferior frontal gyrus, and the inferior parietal lobule (Rizzolatti et al., 1996; Grafton et al., 1996). Note that in the case of the human brain one does not usually talk of mirror

neurons, but rather of a *mirror system*, because single-cell recording is not normally used with humans. However, since there is at least one study (Mukamel et al., 2010) in which single-cell recording has been applied to human beings, we can safely talk of mirror neurons also as far as humans are concerned.

This gives me the chance to make another methodological remark. In what follows, I'm going to rely on the idea that significant similarities exist between macaque monkey brain and human brain[3] (Rizzolatti, Fogassi, & Gallese, 2002; Gallese & Lakoff, 2005), so that, even though in the present work I'll mainly be concerned with human beings, I'm going to extensively report experiments carried out on monkeys working on the assumption that their conclusions may be applied to the human brain. Having said this, we should be aware of a series of differences that have already been observed between macaque monkey brain and human brain also as far as the properties of mirror neurons are concerned. For instance, mirror neurons in the macaque monkey brain fire in correlation with transitive actions only, in which an individual interacts with an object. By contrast, mirror neurons in the human brain fire in correlation with a wider range of actions, including intransitive actions—not directed to a target objects, such as speaking—too (Iacoboni et al., 1999; Buccino et al., 2004). Having said this, it's safe enough to assume that results found in the macaque monkey brain can be carried over to the human brain.

---

[3] We will use the results on monkeys as applying to humans for the simple reason that there is enough evidence to support the notion of an analogy—when not a homology—between the monkey and human brain regions we will be discussing (see Rizzolatti, Fogassi, & Gallese, 2002).

(Gallese & Lakoff, 2005, p. 458)

For all that has been said so far, you might be led to think that the visuo-motor congruence of mirror neurons comes down to the fact that these neurons fire in response to the same kind of *movements*, regardless of whether these movements are executed or just observed. That is, you might think that these neurons fire in correlation with a certain kind of trajectory and speed of a given effector, such as the hand or the mouth. But if this was all that mirror neurons were sensitive to, it shouldn't matter for the sake of the activation of, e.g., grasping mirror neurons whether an object is actually being grasped or not, so long as the arm and the hand display the right kinematics and dynamics. However, this turns out not to be the case.

Umiltà and colleagues (2001) carried out a single-cell recording study in the area F5 of two macaque monkey brains. In this experiment, the following four conditions were compared. All of the conditions featured an experimenter standing in front of the monkey, with a metallic frame interposed between the experimenter and the monkey, allowing the latter to see the experimenter's upper body and arms. In the first (*full vision*) condition, the monkey was shown an action directed to an object: the experimenter's hand started from a given fixed position, moved towards the object, grasped it and held it for about one second. The action just described was thoroughly visible from beginning to end, i.e. since movement onset up until the object had actually been grasped. The second (*hidden*) condition was analogous to the first condition, except for the following detail: an occluding screen was placed so as to occupy the rightmost half of the space enclosed by the metallic frame, in such a way that the second half of the action, in which the experimenter's hand interacted with the object, was hidden from the view of the monkey. Crucially, at the beginning of each trial, before action onset, the occluding

screen was briefly lifted so as to let the monkey see the object placed behind it. The third (*miming in full vision*) and fourth (*hidden miming*) conditions mirrored the first and second one respectively, but with the following variation: no object was present, so in both conditions the experimenter mimed the grasping of an object. In the *miming in full vision* condition, the miming was visible to the monkey from beginning to end, whereas in the *hidden miming* condition part of the miming took place behind the screen.

The results show the existence of mirror neurons that respond selectively to grasping actions, even when the second part of the action—the one in which the object is grasped, which is crucial for the picking out the action as one of grasping—is hidden from view. The same neurons show no response in the miming in full vision as well as in the hidden miming condition. For our purposes, the reported study matters for the following reasons. If you thought that mirror neurons were sensitive to a certain kind of bodily movements, i.e. to a certain trajectory and speed of the arm and hand of the experimenter, this study offers an example of a case in which the same movements performed with vs. without a target object elicit or fail to elicit (respectively) a certain response from a number of mirror neurons. Now, at this point you may want to question that, when the target object is absent, the experimenter's movements are exactly the same as those in which the target object is actually grasped. To rule out the possibility that differences in bodily movements across conditions could account for the results, hand movements were analyzed. No systematic differences connected with the presence or absence of the object were found. In particular, for some of the recorded neurons, hand movements were extremely similar across conditions. For some other recorded neurons, subtle differences could be found in hand

movements between conditions, but no systematic connection existed between differences in hand movements and different responses (Umiltà et al., 2001, p. 160). Note that the considerations I've made so far do not rely on the two hidden conditions[4]. The significance of the latter is going to be discussed at a later stage.

In a nutshell, what's especially important for our purposes is that mirror neurons have been found which do not fire in the absence of an object, even if a grasping action is being mimed by means of the performance of the appropriate sequence of movements. In order to elicit a response in these neurons, an actual act of grasping—one in which an object is actually grasped—has to be carried out, while the relevant bodily movements alone will not suffice to elicit a response[5]. A series of additional results, to be reviewed in what follows, will point us further away from the idea that mirror neuron activation might correlate with mere movements.

2. Outcomes: more than mere movements

Further flexibility in the responses of some F5 neurons was found by means of an experiment by Umiltà and co-workers (2008), in which monkeys were trained to grasp a piece of food with two different kinds of pliers: normal pliers and reverse pliers. The former require the subject to first open and then close one's hand to grasp an object. The latter require the opposite sequence of movements, namely first closing and then opening of the hand. With normal pliers, a typical recorded

[4] Indeed, the same conclusions could have been drawn from Gallese et al. (1996, reported in Csibra, 2007), where it is shown that the sight of a hand pretending to grasp in the absence of a target object fails to elicit a mirror neuron response in the macaque monkey brain.

[5] Some might worry that it is the presence or absence of an object that primarily or even exclusively plays a role in eliciting a mirror neuron response. This worry is going to be addressed in section 4 of the present chapter.

neuron started to fire as soon as the hand started to close, reaching the peak of activation approximately when the piece food was grasped; with the reverse pliers the same neuron would start to fire in register with the beginning of hand opening, and also reached its maximum activation when the food was grasped. The results show that the recorded neurons exhibited the same firing pattern during the grasping phase, regardless of whether this involved normal or reverse pliers—that is, regardless of whether this involved first opening and then closing one's hand as opposed to first closing and then opening. This finding gives further plausibility to the idea that mirror neurons fire in correlation to something that's more abstract than mere movements, since opposite movements can give rise to the same pattern of activation. But what is it that mirror neuron activation correlates with, which is dependent on and yet more abstract than mere movements?

An action can be analyzed as composed of an outcome (or goal, or end) and the means to achieve it. Consider as an example (Csibra 2007, p. 438) the action of sealing an envelope by licking it. Here we have an action whose outcome, sealing the envelope, has been attained by licking the envelope, but could have alternatively been achieved by pressing the envelope on a wet sponge. The same distinction may be applied to the grasping of an envelope. If my hands are free, I will normally grasp it with one of my hands, but if my hands are full as I approach the mailbox, and the mailbox is at a suitable height, I may well end up grasping the envelope with my mouth. Either way, the outcome of my action is grasping the envelope, and the means are either my hand or mouth movements.

Any given action may be interpreted as the attempt to bring about a certain outcome via certain means. What counts as the outcome and what count as the means may vary in interesting ways, of which I'm now going to provide examples.

A seminal experiment by Meltzoff (1988, reported in Gergely et al., 2001) on 14-months-olds was set up so that each infant was visually exposed to the following scenario: an adult model sat in front of a table on which there was a box covered by a translucent panel containing a light bulb; the adult model caused the box to light up by lowering the upper part of his body and touching the panel with his forehead. One week later, the infant was made to come back and was presented with the box with which the adult model had carried out his demonstration. The results of this study show that 67% of the infants imitated the action to which they had been exposed a week before. The action in question is not one that they would have spontaneously performed, so this experiment shows that imitative learning is present at 14 months. Meltzoff (1995, p. 509, quoted in Gergely et al., 2001, p. 325) reasons as follows: 14-month-olds do not just recognize a certain outcome—i.e., the lighting of the panel—but also the means with which the adult model has chosen to achieve this outcome, namely by touching the panel with his head. If infants limited themselves to recognizing a certain outcome to be achieved, namely lighting the panel, they wouldn't necessarily try to achieve it by using their head. The fact that they lean forward with their upper body, according to Meltzoff, shows that these infants imitate the means employed, and not just the outcome of the action they've been exposed to.

An interesting variation on Meltzoff's (1988) experiment was carried out by Gergely and colleagues (2001). Two groups of 14-month-old infants were tested, following a schedule analogous to that of the original Meltzoff's experiment: each infant was exposed to an action performed by an adult model, and was brought back one week later. At that point the infant was allowed to act in the same environment in which he'd witnessed the action performed by the adult model, in

her presence. Once again, a box that could be lit up was displayed on a table. The first group, consisting of 13 infants, was tested in the *hands free* condition, in which the infant could see that the model's hands were free, and yet the model did not use them. Just like in Meltzoff's experiment, the model leaned forward with her upper body and lit up the lamp by touching it with her forehead, which she repeated three times. The second group, consisting of 14 infants, was tested in the *hands occupied* condition, in which the model, before acting on the box, asked another experimenter for a blanket, pretending to be cold. At that point, the experimenter gave the model a blanket, and the model covered her shoulders with it and held it with both hands for the rest of the experimental session. Then she performed the lowering of her upper body and lit up the box with her forehead, and repeated this action three times. The two conditions are meant to differ in terms of the constraints that may account for the model's performance of her action in the specific way in which she executes it. In the *hands free* condition, nothing forces the model to use her forehead instead of, more simply and naturally, her hand (or hands) to light up the box. By contrast, in the *hands occupied* condition, the set up may suggest that the model lights up the box by touching it with her forehead, instead of using her hands, because the latter are occupied, since the model is simultaneously holding tightly a blanket wrapped around her. The question was whether the presence or absence of visible constraints would have a bearing on the way in which the infants chose to imitate the model's action.

Indeed, the two conditions produced different responses in the tested infants. In the "Hands free" conditions, 75% of the tested infants imitated the action they'd been exposed to, thus replicating Meltzoff's (1988) results. By

contrast, out of the infants tested in the "Hands occupied" condition, only 27% imitated the action performed by the model by touching the box with their forehead. The majority (73%), instead, touched the box with their hand—an action that is much easier to perform, and thus more rational. Much more than Meltzoff's (1988) original experiment, this one shows a clear separation between outcome and means, and indeed provides an example of the possibility to achieve the same outcome—lighting up the box—by different means—either touching the box with one's forehead, or with one's hands. In Meltzoff's experiment, the difference between outcome and means was only implicit. I've already mentioned Meltzoff's interpretation according to which the tested infants imitated both the outcomes and the means, but another interpretation could be put forward according to which, in Meltzoff's set up as well as in Gergely and colleagues' *hands free* condition, the outcome had to be identified not so much with *lighting up the box*, but rather with *lighting up the box by touching it with one's forehead*. Gergely and colleagues' *hands occupied* condition provides, on the other hand, the example of an explicit separation between an outcome and at least two means for carrying it out. Another experimental result in a similar vein, clearly showing the difference between outcome and means, is due to Bekkering and colleagues (2000, reported in Grafton & Hamilton, 2007, p. 594). This study shows that, when a child is faced with another person grasping her own ear either ipsilaterally or contralaterally to the acting hand, the child shows a tendency to reproduce the outcome (grasping a given ear) rather than the means (the hand with which the grasping is carried out).

In a nutshell, it is possible to perform different actions that have the same outcome by varying the means to achieve it, which may amount to performing significantly different movements (see Butterfill & Sinigaglia, 2012). The

framework of interpretation featuring outcomes and means is especially useful to make sense of a number of experimental results concerning mirror neurons, such as the one involving reverse pliers (Umiltà et al., 2008) reviewed at the beginning of this section. That experiment featured two actions in which the same outcome—grasping an object—is achieved with different means—opening vs. closing one's hand. Some mirror neurons exhibited a pattern of neural activation which, we could plausibly say, correlated with the achievement of the outcome—grasping the object—while abstracting away from the means—the opening or closing movement of the hand. It is interesting to note that, in the same study, some neurons were also registered in the area F1 (Umiltà et al., 2008, p. 2210), and divided into two functional categories. Those belonging to the first category had a pattern of activation that correlated with the progressive achievement of the grasping outcome, whereas those belonging to the second category fired in relation to hand movements (namely depending on whether the hand closed or opened), and thus differentiated between grasping with normal pliers vs. grasping with reverse pliers. We could say that the firing of the neurons belonging to this second category correlated with the *means*, rather than with the *outcome* of the action under consideration.

It is now worth exploring another sense in which it could be said that mirror neuron activation can be said to correlate with the outcome of an action. Let us consider the case of actions that require a series of different steps to be brought to completion. In particular, I'm going to look at two actions: that of eating a piece of food, which requires first grasping the piece of food and then bringing it to the mouth, and that of placing an object into a container, to perform which one first has to grasp the object, and then to put it into a container. In a single-cell

recording study carried out in the inferior parietal lobule (IPL) of the macaque monkey brain (Fogassi et al., 2005), neurons were studied both when monkeys performed the two above described kinds of action, and when they observed the same actions executed by an experimenter. Let us first focus on the results obtained during action execution, rather than during action observation. Most motor IPL neurons whose activation had been shown to correlate with the act of grasping fired differentially according to whether the monkey's act of grasping was followed by eating as opposed to by placing. It might seem appropriate to say that the neurons in question fired in correlation with the action outcome, in the sense of the result that the subject ultimately wants to bring about.

But can this kind of interpretation be put forward? That is, can we suppose that the activation of these grasping neurons correlates with the outcomes *placing* vs. *eating*, or isn't it the case that the differential activation of the neurons in question reflects a preparation to performing one of two very different kinds of *movement*? The latter deflationary interpretation may seem plausible if we consider that the placing condition requires arm abduction, i.e. movement of drawing the arm away from the body, by contrast with the arm movement required to bring the piece of food to the mouth. In order to rule out the hypothesis that the differential discharge of the grasping neurons can be accounted in terms of a different following movement as opposed to a different outcome, a further experimental condition was added in which the monkey placed a piece of food (or an object) into a container located on the monkey's shoulder (Fogassi et al., 2005, p. 663). As a result, the arm trajectory for this second kind of placing was more similar to the arm trajectory for eating. Once this third condition was added, the neural discharge still differentiated between eating and placing, regardless of the

location of the container. For instance, all the neurons whose activation was stronger in the grasping-to-eat condition exhibited a lower activation in both grasping-to-place conditions, included the one in which the container was close to the mouth. The similarity in arm trajectory didn't elicit a similar response. Interestingly, the wrist velocity for grasping-to-place in the original condition (in which the container was away from the monkey's body) was lower than that in the grasping-to-eat condition, whereas the wrist velocity for grasping-to-place in the modified condition (in which the container was on the monkey's shoulder) was higher than that in the grasping-to-eat condition. The reported differences in wrist velocity make it impossible to account for the recorded responses solely in terms of movements, without making reference to action outcomes.

In the course of this section, I've dwelled on the difference between action outcomes and means to achieve those outcomes. On this basis, it has emerged that an action outcome can be more abstract than the movements employed to fulfil it Working on these assumptions, and on the basis of a number of experimental results concerning mirror neurons, I've shown that mirror neuron activity correlates with action outcomes, rather than with movements.

3. The intention interpretation and its attractiveness

3.1 How do we shift from outcomes to intentions? What changes?

Having shown a correlation between action outcomes and mirror neuron activation, I'd now like to introduce an interpretation according to which mirror neuron activation correlates with a certain kind of intentions, which are later going to be termed *motor intentions*. The full development of the notion of motor intention, together with the explanation of its theoretical implications as far as the

connection between intentions and actions is concerned, is going to occupy the rest of this thesis. For the time being, let us simply see how we get to this idea by working on the assumption, developed and argued for in the previous section, that mirror neuron activation correlates with action outcomes.

Recall the study by Gergely and colleagues (2001) that I reported in the previous section. Suppose we were to modify that kind of scenario in such a way that an adult individual was faced with the model in the *hands occupied* condition, and was given the generic instruction to do the same as the model. In this sort of situation, confronted with such an unusual action, in order to comply with the instruction to do the same as the model, the experimental subject may well wonder whether the model's *intention* was simply to *light up the box*, in whatever way (and touching it with the forehead was reckoned to be the best way to achieve the desired outcome given the simultaneous need to hold the blanket in place), or rather to *light up the box using one's forehead*. Indeed, an agent's intention determines what outcome an action is meant to bring about, and thus, in a context of learning by imitation, the agent's intention determines what the relevant outcome is that the learning subject is meant to reproduce. Thus we have an example of a context in which a question about the *outcome* of an action—namely an event or state of affairs[6]—is reasonably turned into a question about the agent's *intention*—namely a mental state. The transition I've just made from a question about outcomes to a question about intentions will be wholly justified under the theoretical viewpoint in chapter 2 (section 3), where it'll be explained that a necessary condition for a mental state to be an intention is that it should represent an action outcome—that which the subject intends to bring about. Can the same

[6] Csibra (2007) suggests that outcomes are best conceived as states of affairs. I disagree, and I'm going to tackle this question in chapter 4 (section 4).

shift—from a question about outcomes to a question about intentions—be made in the interpretation of mirror neuron activation?

At this point, we're working on the assumption that mirror neuron activation correlates with action outcomes. Given that, minimally, intentions are mental states representing action outcomes, it may seem that, by virtue of this, there's also a straightforward correlation between mirror neuron activation and intentions on the other hand. However, as I'm going to show, there are two kinds of action outcomes, and only one of these two kinds of outcome correlates with intentions. As is going to be explained in the course of chapter 2 (and made more precise in chapter 3, sections 3.2-3.3), another minimal requirement on intentions is that their content should be available to introspection. But it's not obviously the case that the content of all action outcomes is introspectable. In chapter 3 (section 1), I'm going to explain that action production seems to rely at least in part on the existence of motor representations, which, roughly, represent an action outcome by depicting the self in action in a way that determines what the pattern of the subject's movements is going to be, and is therefore apt to be transformed into bodily movements (Jeannerod, 1994, 2006). Now, some of these representations specify action outcomes in striking detail, one that may prima facie seem to likely to defy availability to introspection. Consider the phenomenon of *maximum grip aperture* (Jeannerod, 1981, reported in Jeannerod, 2006, p. 5): whenever one reaches towards an object to grasp it, fingers open more than the size of the target object actually requires, but in a way that is proportional to it. It is likely that some motor representations involved in the act of grasping are going to specify in their content the extent to which fingers open.

So, some motor representations represent outcomes with a very detailed kind of content—one that may, for instance, involve the extent to which fingers should open in a grasping act. For action outcomes of that sort, it is unlikely that there exist *intentions* representing them. It is unlikely, for instance, that there should exist intentions whose content involves the extent to which fingers should open in a grasping act. So, the correlation between mirror neuron activation and intentions doesn't straightforwardly follow from the correlation between mirror neuron activation and action outcomes. This is because there exist different outcomes. Some of them are introspectable, and are thus apt to constitute the content of an intention. Some, on the other hand, are unlikely to be introspectable, and, to that extent, they're not apt to constitute the content of an intention.

Lastly, an observation about one consequence of shifting from an interpretation of mirror neuron activation as correlating with outcomes to an interpretation in terms of the correlation between mirror neuron activation and intentions. One may well interpret the experiments by Meltzoff (1988) and Gergely and colleagues (2001) as supporting the idea that, while acting, one is simultaneously fulfilling many outcomes. From this, however, it doesn't straightforwardly follow that one has as many intentions as the outcomes he's fulfilling. For instance, one may think that, at any given time an individual is acting, she's acting under just one intention, which constitutes the *primary reason* of her acting (see Davidson, 1963, and chapter 2, section 2). This issue will be expanded in section 4 of the present chapter and an answer will be proposed in chapter 6.

3.2 What rests on interpreting mirror neuron activation in terms of intentions?

Even before one assesses the likelihood of the proposal that mirror neuron activity correlates with intentions, one may first of all wonder what rests on interpreting neuron activation in terms of intentions under the theoretical viewpoint. As part of an answer to this question, it is important to notice that a few theorists have treated the hypothesis that mirror neuron activity might correlate with intentions in conjunction with a hypothesis about action understanding and mindreading[7,8]. More generally, some theorists (e.g., Rizzolatti et al., 2001; Gallese & Goldman, 2008; Goldman, 2009*a*) have considered the hypothesis that action mirroring might play a role in action understanding. I am going to suggest that, while there are deep and difficult questions as to what role, if any, action mirroring plays in action understanding, we stand a better chance of answering those questions if we assume that mirror neuron activity correlates with intentions.

While giving a detailed account of the connection between mirror neurons and mindreading in the present work would take me too far afield, it is worth reviewing some of what has been said about mirror neurons and intentions in connection with mindreading, so as to give an idea of the way in which my claim concerning the connection between mirror neuron activation and intentions would be helpful for the purposes of the more general debate concerning action

[7] I'm going to give a detailed explanation of both notions in the course of the following sections.

[8] It is to be specified, however, that not all those who have supported the idea of a contribution of mirror neuron activation to action understanding have subscribed to the idea that mirror neuron activation correlates with intentions. Some have settled for weaker claims on what mirror neuron activation correlates with, even within the context of a defence of a role of mirror neuron activation in action understanding. This idea will be explored in section 3.2.4 in this chapter.

understanding. Later on in this chapter (section 3.2.5), I'm going to sketch an account of the way in which conceiving of mirror neuron activity in terms of intentions might be theoretically fruitful within the context of a version of the Simulation Theory of mindreading, whose tenets I'll expound in due course. Let's see what this is all about.

3.2.1 What is mindreading? How does it connect with action understanding?

Mindreading is the ability to ascribe mental states, such as beliefs, desires and intentions, to another individual (Baron-Cohen, 1995). It is the sort of ability that is required when we want to understand in what mental state a given person is at a given moment, in the light of both her other mental states that are known to us (e.g., beliefs, desires) and the circumstances she's in. In the light of the definition of mindreading given above, in order to be said to mindread, an individual has to fulfil at least two conditions (Goldman, 2008, p. 312). First, one has to judge another person as being in a certain kind of *mental state*, which makes it the case that, for instance, judging that a person is performing a certain action by itself isn't an act of mindreading, and likewise judging that a person is uttering a certain sound isn't an act of mindreading: none of them consists in a judgement about an individual's mental state. Secondly, one has to *represent another individual* as undergoing a certain mental state. Cases of empathizing, so long as the latter is conceived merely as echoing, or resonating with, another's mental state, are therefore not to be considered acts of mindreading insofar as, in order to be said to mindread, it is not sufficient for an individual to represent a certain mental state,

such as an emotion: one must also represent a given personas being in that mental state[9].

What about the connection between mindreading and action understanding? Action understanding, as mentioned at the outset of this chapter, is a process whereby an individual understands what another is doing. The two are connected in interesting ways. We could think that there are, first of all, inferential connections of the following sort. Suppose I see you picking cherries from a tree. If I ascribe to you the belief that cherries are a delicious kind of food, I may understand your action as one of picking cherries in order to eat them. If, on the other hand, I ascribe to you the belief that trees would look a lot better without red fruits hanging from them, I might further ascribe to you a desire, or intention, to free the tree in front of you from a series of ugly red fruits, and understand your action as an attempt to fulfil the desire or intention to do so by picking cherries. Both cases instantiate processes of action understanding that are based on acts of mindreading: a judgement (or more judgements) about another person being in a certain mental state is given, and subsequently that person is understood as performing a certain action or other on the basis of the ascribed mental state.

But it is possible to think that there are also constitutive connections between action understanding and mindreading. Consider again Goldman's (2008, p. 312) example according to which judging that a person is performing a certain action isn't an act of mindreading. Suppose we judged that a person is performing an intentional action, and suppose we further thought that, whenever one

[9] The person in question may well be the subject who is performing the act of mindreading. In this case, we'll say that the act of mindreading is of *first-person mindreading*. Otherwise, the act of mindreading is going to be termed a case of *third-person mindreading*.

performs an intentional action, one thereby has the intention to so act[10]. Then we'd have the grounds for performing an act of mindreading by means of which we also ascribe to the person in question an intention to execute the action that we've judged she's executing. So, in that case, while it would be strictly speaking true that judging that a person is performing a certain action isn't by itself an act of mindreading, this judgement would however license another judgement from the point of view of the subject to the effect that the person in question is in a certain mental state, namely has an intention to so act—and so judging indeed constitutes an act of mindreading.

3.2.2 Simulation Theory and simulation-based action understanding

Having said a few words about mindreading, action understanding, and how they are related we still have to find out how they connect with mirror neurons. The connection goes via one of the leading theories of mindreading, namely the Simulation Theory (Goldman 1989, 1992, 2006; Gordon 1986, 1995)[11]. According to the Simulation Theory, our ability to form judgements about other people's mental states is grounded in our ability to *simulate* those mental states, i.e., in our being able to reproduce other people's mental states and processes in our own mind. More generally, according to the Simulation Theory, mindreaders gain insight into the mental processes of others by using their own ones as a model (see

[10] This is a formulation of what Bratman (1984, 1987) calls the Simple View on the relation between intentions and intentional actions.

[11] The main competitor of Simulation Theory is the so-called Theory Theory (Baron-Cohen 1995; Gopnik & Meltzoff, 1997; Gopnik & Wellman, 1992; Leslie, 1994), according to which we attribute mental states on the basis of a theory about how mental states are connected to behaviour. Lately, theories of mindreading as a whole, namely Simulation Theory and Theory Theory, are facing another opponent as to how social cognition—namely the ability to understand and interact with other people—has appeared: Interaction Theory (e.g., Gallagher, 2001, 2004; De Jaegher, 2009), which does without mental state ascription.

Gallese & Goldman, 1998, p. 496). Let us try to get a rough idea of what this could signify by means of an example. Kahneman and Tversky (1982, reported in Gallese & Goldman, 1998, p. 496) exposed subjects to the following narration: two individuals travelling towards an airport by sharing a car are caught in a traffic jam and thus end up missing their respective flights, which were scheduled for half an hour before their actual arrival to the airport. Mr A finds out that his plane left on time, whereas Mr B gathers that his flight had been delayed, and so he's just missed it by five minutes. Interrogated on which of the two travellers must be more upset, the vast majority of experimental subjects replied that it is Mr B. According to the Simulation Theory, people arrive at the judgement that Mr B is more upset than Mr A by putting themselves in the travellers' shoes and imagining how they would feel in that situation (Goldman 1989, reported in Gallese & Goldman, 1998, p. 496).

In particular, in the given example, subjects are presented with a narration that gives them some information about the travellers' *mental states* while they're en route to the airport (e.g., presumably, their desire, if not intention, to catch their flight, the belief that their flight leaves at a certain hour, etc.). As the narration proceeds, experimental subjects acquire more information about the *circumstances* of the travellers: in particular, about the fact that they get to the airport half an hour later with respect to their scheduled departure. Also, they acquire more information about the travellers' beliefs, as they are informed of their actual flight departure times. At this point, when asked to judge who's feeling more upset, what would experimental subjects do according to the Simulation Theory? First of all, they'd have to simulate their targets' mental states, i.e. reproduce in their own minds mental states that are relevantly similar to those that they presume their

target subject—say, Mr B—to undergo[12]. These are going to be *pretend* beliefs, desires, and so on: experimental subjects know that they're not themselves en route to an airport and that they do not have a scheduled flight, but pretend to believe that they are and that they do. Once that is done, they run the sort of mental processing that they would actually undergo first when acquiring the belief that they've got to the airport half an hour late, and then on learning that their flight left five minutes before, and not half an hour before. What sort of mental state would their own mental processing of all this information output? By letting this processing run on their complex of pretend beliefs, desires, and so on, they obtain as an output a complex of pretend mental states, such as a pretend belief along the lines of "I could have just about made that flight", and the accompanying pretend emotions (e.g., regret).

If the process were to stop at this stage, we wouldn't have an act of mindreading, since, as mentioned beforehand about empathizing, simply replicating someone else's mental state isn't enough to be said to mindread. In addition to the simulation of the target's mental state, there has to be a further step in which the simulated mental state is ascribed to, or, in technical terms, is *projected* onto the target subject (Goldman, 2006, p. 40). At this point, the experimental subjects project all they've obtained as the output of their mental processing onto their target—Mr B, in this case. By applying an analogous procedure to Mr A, experimental subjects obtain as output another complex of pretend mental states that they can project onto Mr A, and, to comply with the

[12] Notice that, for a subject to be said to simulate a target's mental state, there must be the relevant sort of connection between the simulated mental state and the target mental state. If two individuals happen to be in the same mental state, one independently of the other, it cannot be said that one is simulating the other's mental state. Rather, it has to be the case that the simulated mental state is *causally connected* to the target mental state. That is, the target mental state must be the cause of the simulated mental state.

instructions that they've been given, they can compare the two complexes of pretend mental states and, on that basis, judge that Mr B must be more upset than Mr A.

3.2.3 What is the relation between simulation and mindreading?

The above example should have made it clear that Simulation Theory characterizes mindreading as a process during which the mindreader at least tries to replicate the workings of the target's mind (Goldman, 2000, reported in Gallese & Goldman, 1998, p. 497).

> In the simulation scenario there is a distinctive matching or 'correspondence' between the mental activity of the simulator and the target. […] Thus, ST hypothesizes that a significant portion of mind-reading episodes involves the process of mimicking (or trying to mimic) the mental activity of the target agent.
> (Gallese & Goldman, 1998, p. 497)

Now think of a situation in which one individual is performing a certain kind of action that mirror neurons are typically sensitive to—say, she's grasping an apple—while someone else observes her. In the agent, the mirror neurons whose activation correlates with the act of grasping are going to fire. In the observer, the mirror neurons whose activation correlates with the act of grasping are also going to fire—and this activation in observation phase, which we may refer to as *action mirroring*, is caused by the observation of the action of the target. There's a sense in which mirror neuron activation in the observer is a way in which the workings of the observer's mind mimic the mental activity of the target agent. Gallese and

Goldman (1998, pp. 497-498) take this idea even further and claim that if we interpret mirror neuron activation in the agent's brain as a plan to perform a certain action, such as grasping an apple, then, when mirror neurons correlating with that action fire in the observer's brain, they still constitute a plan to perform that action. The mirror neuron activation in the observer's brain is thus treated as alike to the formation of a pretend mental state. Notice that there are grounds for saying that this is the equivalent of a pretend mental state insofar as, as Gallese and Goldman (1998, p. 497) point out, no movement production follows from action mirroring, i.e. from mirror neuron activation in the observer's brain. Furthermore, since the observer has a visual knowledge of the fact that the observed agent is executing the action in question, it is possible to assume that the observer, so to speak, labels the plan constituted by her own mirror neuron activation as belonging to the target. This is the equivalent of the projection stage.

> Thus MN activity seems to be nature's way of getting the observer into the same 'mental shoes' as the target – exactly what the conjectured simulation heuristic aims to do.
> (Gallese & Goldman, 1998, p. 497)

Now, having pointed out a few similarities between action mirroring and simulation, it might also seem obvious to you that there are relevant differences. For instance, I've depicted simulation processes as voluntarily initiated and carried out explicitly by a subject[13], none of which would prima facie seem to be true of

[13] Although it is true that, in everyday life, we may well undergo simulation processes that we do not voluntarily initiate, but are rather elicited by the circumstances, and which are not carried out explicitly by the subject, as when we spontaneously and uncontrollably feel sorry for someone who, we hear, has just been the victim of a catastrophe.

mirror neuron activation. Intuitive as these differences may seem, it's especially hard to pinpoint what they're about. Goldman initially (2006) spelled these differences out in terms of *low-level* vs. *high-level* simulation, and then (Goldman, 2009*b*), in response to de Vignemont (2009), abandoned these terms in favour of *mirror-based* simulation vs. simulation based on what he terms *enactment imagination*.

3.2.4 What is the role of mirroring in action understanding? Causal vs. constitutive role

Either by taking into account both similarities and differences between action mirroring and simulation, or by stressing the similarities between the two, a few theorists have framed action mirroring in terms of action understanding, sometimes in connection with an interpretation of mirror neuron activation as correlating with intentions. I'm going to review a couple of suggested views on action mirroring and action understanding—the first having it that action mirroring constitutes action understanding, while mirror neuron activation doesn't correlate with intentions; the second maintaining that action mirroring contributes to action understanding, and mirror neuron activation correlates with a certain kind of intentions.

Let us start with the mirror neuron researchers themselves (Rizzolatti et al., 2001; Gallese et al., 2004), who have proposed the idea mirror neuron activation in the observer *constitutes* action understanding. Interestingly for my purposes, this proposal comes with an explicit rejection of the idea that mirror neuron activation may correlate with an intention—by stark contrast with what I'm going to suggest later on in this chapter.

> […] the 'direct-matching hypothesis' […] holds that we understand actions when we map the visual representation of the observed action onto our motor representation of the same action. According to this view, an action is understood when its observation causes the motor system of the observer to 'resonate'. So, when we observe a hand grasping an apple, the same population of neurons that control the execution of grasping movements becomes active in the observer's motor areas. By this approach, the 'motor knowledge' of the observer is used to understand the observed action. In other words, we understand an action because the motor representation of that action is activated in our brain.
>
> (Rizzolatti et al., 2001, p. 661)

I'm characterizing the direct-matching hypothesis as making a claim about a constitutive role of action mirroring in action understanding insofar as this seems to be the best way to make sense of a statement such as "an action is understood when its observation causes the motor system of the observer to 'resonate'." It is true that, at the same time, a statement that occurs immediately afterwards to the effect that "we understand an action *because* the motor representation of that action is activated in our brain" (my emphasis) sounds more like indicating that mirror neuron activation in the observer's brain gives a causal contribution to action understanding[14]. Still, I think there are independent grounds on the basis of which we can say that the mirror neuron researchers are making a constitutive claim—for instance, insofar as they appeal to action mirroring as providing "a

[14] As a matter of fact, Goldman (2009, p. 238), more cautiously than me, has claimed that "members of the Parma team *probably lean toward* the constitution construal" (my emphasis).

direct experiential grasp of the mind of others" (Gallese et al., 2004, p. 396, quoted in Goldman, 2009*a*, p. 238).

Now, unfortunately, Rizzolatti and colleagues are not very specific as to how action mirroring by itself should *constitute* action understanding. One of the ways in which they make this connection explicit is by claiming that "[the] observer understands the action because he knows its outcome when he does it" (Gallese et al., 2004, p.). As in this claim, their explanations of the relation between action mirroring and action understanding resort to the notion of an understanding *from the inside*, whereby the observer grasps the outcome of the action she observes by understanding what it would mean for her to perform it[15]—an idea that, since it shifts the focus from the observer agent onto the observer, would be best suited to a framework in which action mirroring is meant to provide a causal contribution to action understanding, and this understanding from the inside is followed by a step that takes care of the attribution of the understood action to the target. We can still make sense of this idea of understanding from the inside by surmising that the mirror neuron researchers' interpretation of action mirroring is only concerned with action categorization, and not with intention attribution. As a matter of fact, the mirror neurons researchers (e.g., Rizzolatti et al., 2001, p. 667) are rather explicit in separating the understanding of an action, which, in their opinion, is provided by action mirroring through direct matching, from intention understanding, which mirror neuron activity by itself, in their opinion, doesn't

[15] […]the parieto-frontal mechanism allows an individual to understand the actions of another individual 'from the inside' and gives the observing individual a first-person grasp of the motor goals and intentions of another individual.
(Rizzolatti & Sinigaglia, 2010, p. 264)

seem to be able to provide. Notice how they interpret the earlier described experiment by Umiltà and colleagues (2001).

> It is important to stress that we are not claiming that, in the experiment of Umiltà *et al.* […], the monkeys understood the *intention* of the agent of the action (that is, *why* the observed action was performed), but only that they understood the *action meaning* (that is, what the agent *did*).
> (Rizzolatti et al., 2001, p. 667, my emphasis)

Let us now move onto a different approach to action mirroring, namely that provided by Jacob and Jeannerod (2005). Jacob and Jeannerod endorse, if only for the sake of argument[16], a version of what they call *motor simulation*, namely the idea that one may represent an action outcome, as well as the observed subject's motor intention, by matching the observed movements onto her own motor repertoire (Jacob & Jeannerod, 2005, p. 22), and thus simulating an action in one's own motor system. An individual's motor repertoire (Rizzolatti & Sinigaglia, 2008) is composed by a series of elements, each of which is a population of neurons that is sensitive to a given action at a certain level of generality (e.g., grasping, or grasping with the hand, or grasping with the hand with a precision grip), and corresponds to the actions that the agent is actually able to perform.

[16] Their ultimate goal is to show that, even if the notion of motor simulation is tenable, it doesn't play either a necessary of a sufficient role in action understanding, insofar as there are actions that may be understood without relying on the simulation of the observed movements (Jacob & Jeannerod, 2005, p. 23), and motor simulation yields understanding of motor intentions only, but not of prior, social or communicative intentions (Jacob & Jeannerod, 2005, pp. 22-23).

They draw a distinction between motor intentions[17] and other kinds of intentions (prior, social and communicative intentions—see Jacob & Jeannerod, 2005, pp. 22-23). Their characterization of motor intentions is twofold. On the one hand, they resort to the idea that an action consists in a goal-directed sequence of bodily movements, and they suggest that this sequence of movements is initiated and monitored by a motor intention (Jacob & Jeannerod, 2005, p. 22). On the other hand, they appeal to the notions of basic action and non-basic action, which they characterize by means of the following example: turning on the light is a non-basic action which can be performed by the basic action of pressing a switch (Jacob & Jeannerod, 2005, p. 22). They also resort to the distinction traced by Searle (1983) between prior intentions and intentions in action[18], and they identify motor intentions with intentions in action, as well as with intentions to execute a basic action[19]. In line with the example they've given about the contrast between basic actions and non-basic actions, they suggest that, on the basis of the prior intention to perform the non-basic action of turning on the light, one may then form the motor intention to perform the basic action of, e.g., pressing the switch with one's right index finger (Jacob & Jeannerod, 2005, p. 22). At this point, Jacob and Jeannerod claim that

[17] I'm going to spend a lot of time on the notion of motor intention from chapter 3 onwards, and I'll show in what sense my characterization of this notion differs from Jacob and Jeannerod's.

[18] In chapter 5, I'm going to show that the distinction between prior intentions and intentions in action is inadequately motivated, and I'll thus suggest dropping the notion of prior intention altogether. I'm presently going to accept talk of prior intentions, as well as of motor intentions in a way that's different from my characterization, for the sake of argument.

[19] I'll broach the topic of basic actions in chapter 6, and there I'll give reasons why the identification between motor intentions—conceived in the way that is going to be formulated throughout this thesis—and intentions to fulfil basic actions shouldn't be made. For the time being, I can anticipate that one of the reasons for not identifying the two notions is that there are more motor intentions than intentions fulfilling basic actions.

> by simulating the agent's perceived movement of pressing the switch with his right index finger, an observer will understand the agent's motor intention to execute the basic action, not his prior intention to execute the non-basic action.
> (Jacob & Jeannerod, 2005, p. 22)

Jacob and Jeannerod therefore outline a view of action mirroring such that:

(a) mirror neuron activity in the observer's brain correlates with intentions—a specific kind of intentions, as I'm also going to argue—and
(b) this correlation serves the purposes of action understanding through what they refer to as *motor simulation*, with the proviso that action understanding mediated by motor simulation is only going to grant access to motor intentions, and not to other kinds of intentions.

Let me take stock. While, as I said, the mirror neuron researchers are not overly explicit in spelling out how action mirroring could constitute action understanding, there's a more general thought, explored in the previous section, according to which action mirroring can be compared with the simulation stage of a simulation process. If that is so, it becomes plausible to suppose that action mirroring needs to be supplemented by an equivalent of the projection stage which is part of a process of simulation[20]. So, the idea that action mirroring might play a *causal* role, rather than a *constitutive* one, in action understanding seems to stand a better

[20] In the previous section, it was reported that Gallese and Goldman (1998) suggest that the projection stage might be obtained through the observer's visual knowledge of the fact that the observed agent is executing the action in question: by means of that knowledge, the observer would label the plan constituted by her own mirror neuron activation as belonging to the target.

chance to be made to work. Having said this, there's still a lot of work to do in order to clarify how mirroring-based action understanding could take place. I claim that a proper interpretation of what mirror neuron activation correlates with is going to bring us closer to a solution of the issue as to how mirroring-based action understanding works. After all, if mirror neurons are to play any role at all in the process of action understanding, we need to be clear on just what they contribute to this process. I claim that, in particular, interpreting mirror neuron activation in terms of intentions is going to be a helpful step towards the comprehension of mirroring-based action understanding, insofar as being provided with the intention that the observed agent is trying to fulfil is fundamental for understanding what she's up to. In the following paragraph I'm going to sketch a version of Simulation Theory applied to action understanding that acknowledges a causal role to action mirroring and that works on the assumption that mirror neuron activity correlates with intentions of a specific sort.

3.2.5 How the intention interpretation could help identifying the role of action mirroring in action understanding

Having explored the general connection that there can be between interpreting mirror neurons in terms of intentions and theories on how action understanding takes place, I'm going to review two experiments to give an example of why interpreting mirror neuron activation in terms of intention might be fruitful under the theoretical viewpoint, first independently of, and then in conjunction with, a version of the Simulation Theory of mindreading applied to action understanding. I'll leave the rest of the explanation of the theoretical

advantages of the hypothesis concerning the correlation between mirror neuron activation and intentions to chapter 6 (section 5).

Let me start with a study conducted by Ferrari and colleagues (2005, reported in Csibra, 2007, pp. 445-446) on macaque monkeys. Monkeys underwent a visual training of actions that required tools, such using sticks to pick up pieces of food. This visual training consisted in letting monkeys passively observe the experimenters using the tools in question. Crucially, monkeys never learned how to use these tools themselves. In support of the fact that these monkeys were unable to use the tools involved in the actions shown, it is reported that, for instance, one of the tested monkeys, when faced with a piece of food outside its cage, did not even make an attempt to use a stick to get hold of it. Still, after two months of observational training, some mirror neurons that would normally discharge when monkeys themselves grasped a piece of food with their hand fired in response to the sight of an experimenter picking up a piece of food by means of a stick. Ferrari and colleagues (2005, p. 216, reported in Csibra, 2007, p. 445) interpret their finding in terms of a similarity between the *goal* (or, we may say, the *outcome*) of the observed action on the one hand and that of the action that the monkey itself would be able to perform in a similar situation on the other hand. So, the monkey mirror neuron response to the observation of an action that the monkey was unable to execute corresponded to an action with the same outcome as that of the observed agent, but an action that the monkey was able to perform.

These results can be interpreted in the following terms. Grasping a piece of food is an outcome that can be achieved, among other ways, by using one's hand (or hands) or by using sticks. One may thus generically intend to grasp a piece of food, or specifically intend to grasp a piece of food by using sticks—this will be the

case for the experimenter who's responsible for the visual training of the monkeys—or intend to grasp a piece of food using one's own hand. The experimenter's action of grasping a piece of food is understood by the monkey not just as fulfilling the *intention to grasp a piece of food*, but as fulfilling the specific intention that the monkey itself would be able to form on the basis of its own abilities, which enable it to grasp a piece of food with its hands. It would seem, then, that the experimenter's intention is mapped onto the intention relevant to the circumstances at hand that the monkey would be able to form in order to achieve the outcome of grasping a piece of food.

Gazzola et al. (2007) have conducted the following study on aplasics, namely individuals born without some limbs. In the study of Gazzola and colleagues, two aplasic individuals born without hands or arms underwent an fMRI scan while they watched hand actions, and their brain activity was studied in comparison with that of sixteen typically developed individuals. During this study, both in aplasics and in typically developed subjects, effector-specific areas of the brain were individuated as they executed actions with different effectors. The main finding of this study is that both aplasic individuals, while observing hand actions, activated those areas in their brain—part of the mirror neuron system—that had previously been shown to be involved in the execution of foot or mouth actions. Just like in the study by Ferrari and colleagues (2005), here we have an outcome—a given action—that can be executed by different means—hands, mouth or feet. A plausible interpretation of this finding would be that aplasics recognize the typically developed individuals' intention to execute hand actions, and map it onto the intention that they would form in order to perform the same action, an intention which is going to specify the effectors that are actually available to them,

namely mouth and feet[21]. That is, they re-interpret the observed action in the light of their own motor repertoire.

The possible interpretation of the two above reviewed studies has an intrinsic plausibility, to the extent that it builds up on an interpretation based on action outcomes that the previous sections should have shown to be more appropriate with respect to an interpretation based solely on movement parameters. This is not the whole story, however, insofar as the proposed interpretation would neatly fit into a bigger theory concerning action understanding, namely a version of Simulation Theory, which is faithful to the fundamentals of the theory as I've expounded them: mirror neuron activation in the observer reflects the intentions that the observer, given her possibility for action (we might say, given her motor repertoire), would form to achieve the outcomes that she sees the target agent achieving on the basis of the latter's possibilities for action. In other words, by means of action mirroring, an observer literally puts herself in the target's shoes, and has a grasp of what it would be *for her* to execute the action that the target is performing.

3.3 Some provisional conclusions

Let me sum up where we've got so far. In the present section, I've put forward an interpretation of a series of experimental results concerning mirror neurons in terms of their correlating with:

(a) the intention of the acting agent, when mirror neurons fire during action execution;

[21] See chapter 3 for an alternative interpretation (in terms of action representations) of the same study compatible with the present one.

(b) the intention of the observed agent, mapped onto the intention that the observing agent would form in the circumstances in which the observed agent acts, when mirror neurons fire during action observation (Ferrari et al., 2005; Gazzola et al., 2007).

The interpretation of mirror neuron activation in terms of intentions has an intrinsic plausibility that builds on the unavailability of a deflationary interpretation that relies solely on movement details, and on the likelihood of an interpretation in terms of action outcomes. Still, there is a non-trivial step involved in shifting from an interpretation in terms of action outcomes to one in terms of intentions, insofar as

(i) it is not clear that all outcomes that one fulfils while acting are apt to constitute the content of an intention, since the content of an intention should be introspectable, and

(ii) working on the assumption that, while acting, one simultaneously fulfils many outcomes, it is not obvious that one thereby has as many intentions as outcomes she's fulfilling.

An interpretation in terms of correlation between mirror neuron activation and intentions would also lay the foundations for a version of the Simulation Theory of mindreading applied to action understanding.

4. Which intentions? How many intentions?

For the sake of argument, let us suppose for the time being that an interpretation of mirror neuron activation having it that mirror neuron activation correlates with intentions could be made to work. A lot more still needs to be specified about what this interpretation would have to say about the various experimental results reported in this chapter[22]. For instance, consider once again the study by Ferrari and colleagues (2005) and that by Gazzola and colleagues (2007). As I've said in the previous section, both can be seen as featuring an outcome that can be achieved by multiple means. Take grasping a piece of food, which can be achieved, among other ways, by picking it up using sticks or with one's hands. Even assuming that we're going to employ the notion of intentions to characterize the outcome which a given subject tries to achieve, how many intentions, and which intentions, should we say are in play? Would it be more appropriate to say that the experimenter has the more general intention to grasp a piece of food, or the more specific intention to grasp a piece of food by means of sticks, or that it has two intentions—a more general one to grasp a piece of food and a more specific one to grasp a piece of food by using sticks? Likewise, in Gazzola et al. (2007), should it be said that the observed individual has both the intention to perform a given action tout-court, or a given action with a specific effector?

Now recall the study by Fogassi and colleagues (2005), as it poses a problem analogous to the one just raised. The authors suggest an interpretation of their findings according to which some of the recorded mirror neurons "not only code the observed motor act but also allow the observer to understand the agent's intentions" (Fogassi et al., 2005, p. 662), from which it would seem that they

[22] A re-interpretation of these results in the light of motor intentions is going to be provided in chapter 6 (section 5).

conceive of grasping and eating as the intentions of the acting individual, whereas it's not clear that the act of grasping has an intention associated to it. Alternatively, it would be possible to say that:

(i) the agent intends to grasp in order to place (only one intention is involved), or

(ii) the agents intends to grasp, and also intends to grasp in order to place, or

(iii) the agent intends to grasp and also intends to (subsequently) place, or

(iv) the agent intends to grasp, and also intends to subsequently place, and also intends to grasp in order to place.

Notice that the equivalent of (iii) or (iv) weren't available for the previously mentioned studies (Ferrari et al., 2005; Gazzola et al., 2007), since they didn't feature actions that could be segmented into temporally distinct components such as *grasp* plus *eat* or *grasp* plus *place*. A proposal to interpret experimental results concerning mirror neurons by itself doesn't specifically commit us to any of the earlier proposed alternatives. If the suggested interpretation in terms of intention can be made to stand, the issue as to which intentions and how many intentions are involved in the performance of a given action needs to be dealt with[23].

Before I proceed any further, it is worth stressing what has been done and what hasn't been done so far. I've started out by considering an interpretation of mirror neuron activation according to which the latter correlates with movement

[23] And, indeed, it is going to be dealt with in chapter 6.

parameters only. This interpretation has been shown to be untenable, given the various examples of mirror neuron activation abstracting away from fine movement details, and an alternative interpretation in terms of action outcomes has been proposed. Building up on the action outcome interpretation, it has been proposed that mirror neuron activation correlates with the intention (or intentions—as I've highlighted throughout the present section) of the agent and the corresponding one (or ones) of the observer. It is essential to note that my rejection of an interpretation of mirror neuron activation in terms of movement parameters shouldn't be read as a complete dismissal of the role of bodily movements in the characterization of the relevant high-level outcome. Unfortunately, the interpretation of mirror neuron activation in terms of movement parameters has sometimes been formulated as antithetical to that in terms of high-level outcomes. The following section is devoted to showing that this contraposition is mistaken.

5. Mirror neuron activation: high or low?

In this section, I'm going to review an argument that has been proposed by Csibra (2007) to the effect that mirror neuron activity can be embedded in one of two competing models of action understanding—direct matching and action reconstruction. Csibra further argues that the latter is more plausible than the former, partly by resorting to a number of experimental results concerning mirror neurons, results that Csibra takes to witness that observed actions are understood at a relatively high level of abstraction. I don't mean to side with Csibra nor am I interested in adjudicating between the two hypotheses that he contrasts concerning how action understanding works. Rather, I'm interested in defending

the idea that there is a correlation between mirror neuron activation and intentions of a specific kind, and this idea is likely to be compatible with either hypothesis on how action understanding works considered by Csibra. Reviewing Csibra's argument is mainly going to serve the purpose of removing obstacles for the idea that I ultimately want to defend, for the following reason. Csibra's argument seems to be based on an assumption on the individuation of action outcomes such that, above a certain level of abstraction, the individuation of action outcomes doesn't rely on movement details at all. I mean to show that this assumption is mistaken, and this is going to provide the basis for arguing in favour of a specific kind of intentions—motor intentions—corresponding to the kind of outcomes—which I'm going to term *motor outcomes*—that are seemingly not contemplated in Csibra's theoretical framework.

Csibra starts by considering that an action may be understood, and therefore reproduced, at different levels of interpretation. Over the previous sections we saw many examples of this: in Meltzoff's (1988) experiment the action in question could be understood either as a lighting up the box or as a lighting up the box using one's forehead, and in Gazzola and colleagues' study (2007) a given action could be characterized in a way that does or does not specify the effector employed, and so on. An influential proposal (Rizzolatti et al., 2001) about the way in which mirror neuron activation might mediate action understanding consists in the so-called *direct-matching hypothesis*, which states that "an action is understood when its observation causes the motor system of the observer to 'resonate'" (Rizzolatti et al., 2001, p. 661, quoted in Csibra, 2007, p. 436). According to this hypothesis, an observer would reproduce the observed action in her own motor system, and would thus be enabled to understand the outcome of the action she

observes[24] (Csibra, 2007, p. 436).

Now, the reproduction of an action in one's motor system is meant to consist in a conversion from the visual description of the observed action given at some level of interpretation into the motor description of the same action, which is produced in the observer's motor system (Csibra, 2007, p. 437). Csibra considers it a crucial question at what level of action interpretation the reproduction of an action in the observer's motor system takes place. He claims that the notions of *direct matching* and *motor resonance* are likely to be based on the idea that the above mentioned conversion must take place at a relatively low level, where movement details are described, and possibly in great detail—and indeed, according to him many results concerning mirror neuron activation seem to be compatible with this idea[25]. At the same time, however, Csibra (2007, p. 443) claims that a series of other results concerning mirror neuron activation seem to indicate that the conversion between the visual description of an observed action into a motor description must take place at a very high level of interpretation. In support of this idea, he mentions results that have been reported earlier in this chapter: mirror neurons do not respond to mimed actions, even though the movement details of the mimed action are very similar (if not the same as) those of a target object-directed action (Gallese et al., 1996). Csibra also quotes the results obtained by Umiltà and colleagues (2001): there is no activation of mirror neurons

[24] The process of action understanding mentioned in the context of this hypothesis has been referred to as motor simulation (e.g., by Jacob & Jeannerod, 2005), and indeed seems to fit the general principles of the Simulation Theory of mindreading.

[25] For instance, he mentions the existence of strictly congruent mirror neurons, whose firing corresponds to relatively fine details of the action that is being performed or observed—that is, these neurons correlate not just with grasping with the hand, but also with the specific kind of grip (precision grip or whole hand prehension) employed (Gallese et al., 1996).

whenever an action is performed partly behind an occluding screen if the monkey knows that no target object has been placed behind the screen, in spite of the fact that the relevant movements are being performed—although the presence of the relevant movements should be enough, according to Csibra, for low-level mirroring to occur.

According to Csibra (2007, pp. 446-447), the direct-matching hypothesis as it is normally presented by its supporters is flawed by a tension between the following conflicting ideas: on the one hand, action mirroring is taken to rely on mechanisms of low-level resonance, sensitive to relatively fine movement details, and, on the other hand, action mirroring would seem to be significantly independent of fine movement details (as in Gallese et al., 1996, and in Umiltà et al., 2001), and rather to reflect action understanding at a high level. To the extent that action mirroring can be shown to consist in a close reproduction of the movement details of the action that is being observed, the less there seems to be reason to identify action mirroring with action understanding—i.e., with the understanding of the high-level outcome that is being pursued by means of the action in question. By contrast, to the extent that action mirroring is shown to rely on an interpretation of actions at a high level, there seems to be no role for action reproduction to play in the observer's motor system at a relatively low level for the purposes of action understanding. In Csibra's own words:

> The tension arises from the fact that the more it seems that mirroring is nothing else but faithful duplication of observed actions, the less evidence it provides for action understanding; and the more mirroring represents high-level interpretation of the observed actions, the less evidence it provides that this interpretation is generated by low-level motor duplication.

(Csibra, 2007, p. 447)

Taking this tension into account, Csibra proposes two alternative models of action understanding, which are primarily differentiated by the level at which action interpretation takes place, and, on the basis of that, assign different roles to mirror neuron activation. In one model, which is still called *direct matching*, the interpretation of the observed action takes place at a relatively low level. Mirror neuron activation provides this low-level motor resonance, and thus contributes to action understanding in the way that is going to be explained in the following paragraph. By contrast, an alternative interpretation—*action reconstruction*—has it that the interpretation of the observed action occurs at a relatively high level, possibly outside the motor system, and is then passed on to the motor system, and it's only at that point that mirror neurons fire.

Csibra makes reference to hierarchical models of motor control (Wolpert et al., 2003), according to which there are hierarchically organized motor modules that are reciprocally connected. On the basis of these models, he links the two proposed alternatives concerning the level at which action interpretation occurs with two hypotheses concerning the propagation direction of activation within the action control system (Csibra, 2007, pp. 441-442). In the direct-matching model, according to which action interpretation takes place at a low-level, motor resonance would determine an upward propagation in the observer's motor system producing an estimate of the higher level subgoals and goals likely to be at the origin of the observed action. On the alternative model, that of action reconstruction, according to which action interpretation occurs at a high level, the high-level outcome is mapped onto the observer's action control system, where it can be propagated downwards thus eliciting the relevant covert emulation of the

action (Csibra, 2007, p. 441). In the former case, it can be said that mirror neuron activation *contributes to* (or even *constitutes*—see discussion in section 3.2.4 of the present chapter) action understanding. In the latter case, it is more appropriately described as *reflecting* action understanding (Csibra, 2007, p. 443).

Csibra's argument can be summarized as follows:

(1) Action reproduction in the observer's motor system can only take place at one level of a hierarchy;

(2) There is a level at which action outcomes encompass movement details, and a level at which action outcomes abstract away from movement details, and the two are different (and are, respectively, low and high);

(3) Mirror neuron activation is action reproduction either at a low or at a high level, but not both.

> One cannot have one's cake and eat it too: the discharge of a set of MNs cannot represent the activation of the observer's motor system at low and high levels at the same time.
> (Csibra, 2007, p. 447)

(4) If mirror neuron activation is action reproduction at a low level, then it *contributes* to action understanding in the way specified by the direct-matching model. If mirror neuron activation is action reproduction at a high level, then it *reflects* action understanding in the way specified by action reconstruction.

Csibra further argues for action understanding as opposed to direct matching. As I

said, I'm not going to pronounce myself on this contrast, but I mean to question assumption (2), according to which the specification of an action outcome should either involve or fail to involve the movement details of an action—the assumption that makes direct matching an alternative incompatible with action reconstruction. Earlier on in this section I mentioned some experimental results (Gallese et al., 1996, and Umiltà et al., 2001) on which Csibra seemingly bases this assumption, but I think the justification for (2) heavily relies on Csibra's idea that, in a series of experiments concerning mirror neurons that I'll discuss below, contextual cues play a predominant and seemingly exclusive role in the individuation of an action outcome, as opposed to movement details. I'm going to show that, in the light of the details of the experiments under consideration, Csibra's reflections can be shown to be defective, and thus his assumption (2) concerning the individuation of action outcomes can be shown to be questionable. I'm going to rely on the correlation between mirror neuron activation and action outcomes established in sections 1 and 2 to point to the existence of action outcomes, which I'm going to term *motor outcomes*, that are at the same time dependent on the movements employed to achieve them and also independent of the fine details of those movements.

Recall the study by Fogassi and colleagues (2005), and let us now look at the results obtained in observation condition. Many motor IPL neurons fired also while the monkey observed the grasping for eating and grasping for placing actions performed by others. Most of them exhibited a differential activation depending on whether the grasping act in correlation with which they responded was going to be followed by eating or placing. Note that the differential discharge in question occurred before the beginning of the following act, namely eating or

placing. It is to be specified that the container would be present only in the placing trials, and not in the eating ones. Csibra makes the following claims.

> Although the actions themselves did not carry information about the further goal of the actor in the studies cited above, the context did. For example, the monkey in the Fogassi *et al.* (2005) study could figure out the further goal of the observed action from the kind of object (food or non-food) involved, and whether or not a container was present.
> (Csibra, 2007, p. 445)

While these considerations make sense in principle, it is also true that Fogassi and colleagues (2005, p. 664) controlled for the effect that the specific kind of object produced on the monkey. A series of neurons (20) were therefore tested in a modified version of the observation task, in which three actions were shown to the monkey: grasping food to eat, grasping food to place, and grasping solid to place. The recorded neurons still exhibited selectivity towards the specific action following grasping, regardless of the object employed. To be precise, 12 of the recorded neurons did not show any difference in activation between grasping food to place and grasping solid to place, while 8 neurons, which were selective for grasping to eat, discharged more strongly in response to grasping food to place as opposed to grasping solid to place. In a nutshell, although the nature of the object involved could in principle have been sufficient to account for a neural response anticipating the outcome of the action, it turned out not to be sufficient.

We haven't yet discussed the role of another contextual cue, namely the presence or absence of the container in Fogassi and colleagues' experiment—a role that Csibra considers prominent, to the detriment of movement details, in the

individuation of a given high-level outcome, as witnessed by this reflection:

> if we accept Fogassi *et al*.'s (2005) argument that the selectivity of MNs was independent of the kinematic parameters and reflected 'intention' understanding based on contextual cues, then nothing in this study provided evidence that such an understanding is based on low-level mirroring (i.e. motor resonance).
>
> (Csibra, 2007, p. 447)

But note that Fogassi and colleagues only establish that their results cannot be accounted for *solely* in terms of movement parameters, insofar as no systematic connection holds between those parameters and mirror neuron responses[26]. But this doesn't amount to denying that movement parameters play any role whatsoever in action understanding. At this point it is especially helpful to consider Umiltà and colleagues' (2001) study insofar as its two conditions featuring an occluding screen are concerned. Conclusions analogous to those that I'm going to draw about that study should hold for the experiments by Fogassi and colleagues. First, a consideration of a more general kind is in order. It cannot be said that only contextual cues, consisting in the visible presence of the relevant objects (the object to be grasped in the study by Umiltà and colleagues and the container in that by Fogassi and colleagues), are responsible for the recorded mirror neuron response. This is because, in the first place, mirror neurons do not respond to the sight of objects alone (Gallese et al., 1996; Rizzolatti et al., 1996; both reported in Umiltà et al., 2001)—by contrast with canonical neurons (Murata

[26] This lack of systematic connection was verified on monkeys in the action execution condition, and it should be reasonable to assume that it holds also for a human experimenter performing the same actions.

et al., 1997). Secondly, mirror neurons do not fire until the beginning of the action, even though the contextual cues are already present (Fogassi et al., 2005). In the study by Umiltà and colleagues (2001, p. 160), seeing the object that is going to be acted upon being placed behind the occluding screen doesn't elicit any mirror neuron response.

The condition in which the occluding screen is employed and an object is acted upon is useful for the following consideration. The neurons that fired in correspondence with the late phase of grasping and during holding produced the maximal response at the end of the hidden action. Furthermore, mirror neuron activation was significantly larger in the phase in which the hand was hidden compared to that in which the hand was still visible in the population as a whole (Umiltà et al., 2001, p. 160). So it is not necessary for a contextual cue to be visible in order to elicit a mirror neuron response. Taken together, these results are not meant to show that contextual cues do not play any role in eliciting the relevant mirror neuron response, but they can be taken to signify that contextual cues do *not* play the predominant or even exclusive role—to the detriment of movement details—that Csibra seems to grant them.

I started out by meaning to question Csibra's underlying assumption on the individuation of action outcomes such that, above a certain level of abstraction, the individuation of action outcomes doesn't rely on movement details at all. This assumption relied on the idea that, whenever mirror neuron activity does not correlate with fine movement details, it can be mainly or even wholly accounted for by invoking the presence of relevant contextual cues. However, my discussion of the experimental results that Csibra appeals to shows that contextual cues alone cannot account for the obtained mirror neuron response, and thus it's not the case

that mirror neuron activation, in those cases, correlates with a high-level outcome whose characterization is completely independent of movement details. It would rather seem, on the basis of the discussion of the role of contextual cues in the experiment by Fogassi and colleagues, that action outcomes exist that are at the same time relatively high level, insofar as they abstract away from fine movement details, while their characterization is still relatively low level, insofar as it essentially relies on movements details, as contextual cues alone cannot account for them. These action outcomes—*motor outcomes*—are going to play a key role in the definition of *motor intentions*, namely the intentions with which mirror neuron activation correlates.

Conclusion: motor outcomes—not so high, not so low

Once it is established that mirror neuron activation correlates with motor outcomes, which, on the one hand, cannot be reduced to mere movements, but at the same time to some extent depend on movements, we can set out to suggest that mirror neuron activation correlates with a kind of intentions that I'm going to term *motor intentions*—in a related but different sense with respect to Jeannerod (1994), Pacherie (2003, 2006, 2008) and Jacob and Jeannerod (2005). The notion of motor intention is what is going to constitute an answer to the central question of the thesis, namely how actions are connected with intentions. A hint to why motor intentions are going to provide this connection can be found in the notion of *motor outcome*, which has been worked out in the previous section. Motor outcomes are at the same time relatively high level, insofar as they abstract away from fine movement details, and low level enough that their characterization essentially relies on movements details. So, if the idea of a motor intention can be

made to work, the sort of outcome that it represent is going to be a motor outcome in the way just characterized, and, in the light of the discussion that has occupied section 3 of the present chapter, it's going to be possible to say that mirror neuron activity correlates with motor intentions. Still, a lot of steps are still required to reach a satisfactory notion of motor intention, and to show that it answers the question as to how intentions connect with actions. First of all, we're going to need to tackle the notion of intention itself, in order to identify the distinguishing characteristics of an intention, so as to be able to show that a motor intention satisfies them. After that, I'm going to focus on the distinguishing characteristic of motor intentions, namely the relevant similarity in content that they bear to motor representations. This is going to be the topic of chapter 3. There the notion of motor intention will be introduced by appealing to the idea that intentions as well as motor representations represent action outcomes. Chapters 3 and 4 are going to deepen the idea that action outcomes represented by motor intentions are not so high, insofar as they are concerned with bodily movements (and, as a matter of fact, specify sequences of bodily configurations), and they're not so low insofar as they're endowed with a certain degree of abstraction with respect to greater detailed representations of movements. The notion of motor intention could in principle be thought to be similar to Searle's (1983) notion of intention in action. Chapter 5 is going to show that, as a matter of fact, Searle's distinction between intentions in action and prior intentions is inadequately motivated, in particular as far as the temporal dimension it involves is concerned. So the notion of motor intention is significantly different from that of intention in action. Once we reach chapter 6, we'll finally be in the position to appreciate that intentions can be ordered hierarchically, and that motor intentions occupy the lowest levels of this

hierarchy. The special connection holding between motor intentions and motor representations, which is going to be spelled out in chapters 3 and 4, is going to explain why motor intentions provide the connection between intentions and actions: since motor representations are the proximate causes of movements (see Pacherie, 2000, p. 403), and motor intentions have a kind of content that is relevantly similar to that of motor representations, motor intentions are the best candidates among intentions for providing a linkage between intentions and actions: as it were, they're not so high, insofar as they bear relevant similarities to motor representations, and they're not so low, since, after all, they are intentions.

§2. What is an intention?

Introduction

> Very often, when a man says 'I am going to do such-and-such', we should say that this was an expression of intention. We also sometimes speak of an action as intentional, and we may also ask with what intention the thing was done.
>
> (Anscombe, 1957, p. 1)

As this quote illustrates, providing an account of what intentions are is made difficult by their appearance in at least three forms: intentions for the future, as when I intend to open the curtains in an hour, intentions with which one acts, as when I open the curtains with the intention let more light into the room[27], and intentional action, as when I intentionally open the curtains. Desirable as it might be (since we could plausibly think that there must be just one concept underlying these three forms of intention), it is far from easy to give a unitary account of intention. Radically different accounts of intentions have ensued depending on

[27] This example appears in Bratman (1987, pp. 128-ff.). There Bratman distinguishes different notions of acting with an intention, beginning with the difference between what he calls 'acting with a *further* intention', of which an example is opening the curtains with the intention of washing the windows, and acting with an intention (tout-court), such as opening the curtains with the intention of *thereby* getting more light. In the first case, the intentional action is included in a larger plan, part of which consists in a distinct and successive action of washing the windows. In the second case, the action of opening the curtains is all that's needed in the given circumstances to obtain more light. These distinctions will be put to work in chapter 6.

which form of intention has been taken as primary. In what follows, I'll attempt to give an account of intentions and intentional actions.

1. What is an intentional action? A methodological claim on relating intentional action to responsibility attribution

First of all, I'm going to consider the idea that the notion of intentional action is bound up with that of responsibility attribution. We may think, that is, that insofar as one acts intentionally, one may then be held responsible, and thus praiseworthy or blameworthy, for what she does. This sort of connection indeed seems to be part of our folk psychological intuitions about the notion of intentional action[28], if you consider the following example: if you step on my foot, and it turns out that you've done that intentionally, I may consider you blameworthy for it and therefore be very angry at you. On the contrary, if you step on my foot and I learn that what you've done is not intentional, then I may be inclined not hold you blameworthy for it[29], and thus forgive you. I'm now going to review some results showing that the connection between intentional action and considerations to do with responsibility, blame or praise attributions should be regarded with suspicion as it's not especially reliable. With this, I don't mean to deny that the notion of intentional action is irrelevant for the purposes of ascribing responsibility, blame

[28] I'll later say to what extent I mean to rely on folk psychology, and why.

[29] Bratman (1987, p. 1) provides a similar example:

> If I poison the inhabitants, but do not do so intentionally and do not act with the intention of poisoning them, you will react very differently to me than if I had acted with that intention and poisoned them intentionally; and so will our legal institutions.

But cf. the example concerning a drunk driver (Mele & Moser, 1994, p. 51) mentioned later on in the chapter.

or praise: on the contrary, as stated at the beginning of the present work, I think that the notions of intention and intentional action have a bearing on responsibility, blame or praise attribution. What I set forth to deny is the converse kind of connection, namely that, to clarify what an intentional action consists in, we should resort to notions of responsibility, blame or praise.

Knobe (2006) has looked into people's intuitions as to whether specific actions should be judged intentional or not. He presented individuals with pairs of cases that are almost exactly alike, except that in one a given behaviour has beneficial consequences towards someone or something, and in the other case the same behaviour brings damage to someone or something. As a result, people's intuitions differ on the two cases. Let us see how by looking at an example of a pair of cases (Knobe, 2006, pp. 205-206).

Here is the first case. The vice-president of a company submits the following proposal to the chairman of the board. The company could start a new program, which is going to increase the profits, but also damage the environment. The chairman agrees to start the new program on the basis of the fact that he wants to increase the profits, while he doesn't care about the environment. The program is therefore started, and the environment is indeed damaged. Question: did the chairman damage the environment intentionally? Most people will answer affirmatively to this question, and, asked to motivate their answer, they'll resort to the fact that the chairman knew in advance that the program would damage the environment, and yet agreed to start it.

Now consider the second case. This is entirely analogous to the first one, except for the fact that the proposed program is going to *help* the environment, rather than damaging it. Once again, the chairman shows indifference towards the

effects of the program on the environment, and agrees to start it. Now, did the chairman *help* the environment intentionally? This time most people are going to say that the chairman does *not* help the environment intentionally. In terms of knowing in advance, though, the second scenario is analogous to the first one in that the chairman is indeed aware of the effects of the program on the environment before starting it.

These results suggest that people's judgements as to whether a given action is intentional or not seem to be influenced by considerations concerning the moral status of the behaviour in question—namely considerations as to whether that behaviour is good or bad. The results are replicated even once the pair of cases is modified so as to involve the bearing of trying, foresight and skill (Knobe, 2006, pp. 208-ff.). This may be taken to mean that what people have in mind when they express judgements about an action being intentional or not reflect rather a question as to whether the piece of behaviour in question is one for which the subject may be held responsible, and, more specifically, blameworthy or praiseworthy (see Adams & Steadman, 2004, mentioned in Knobe, 2006, p. 214[30]; and Mele, 2001).

While I think that these results should be taken to indicate that moral considerations distort intentional action attributions, Knobe instead rationalizes people's reactions to the proposed pairs of cases. He does so by suggesting a view according to which people use the notion of intentional action as a tool for determining how much praise or blame a given person deserves (2006, pp. 225-226). This determination, according to Knobe, goes via a two-step process

[30] "When a person utters the sentence 'He didn't do that intentionally,' there is often a clear implicature that the agent is not to blame for what he has done." (Knobe, 2006, p. 214)

whereby a first sub-process determines whether the given behaviour is good or bad, and, on the basis of that, which features are relevant in order to judge it as intentional, and a second sub-process looks for the relevant features in the behaviour under examination and, on the basis of the presence or absence of those features, outputs an intentional judgement. Knobe, in short, endorses a view according to which moral considerations, rather than distorting intentional action attributions, instead turn them into useful tools for evaluating whether a given behaviour is praiseworthy or blameworthy.

I'm going do draw a different conclusion from Knobe's results. Consider the alternative view according to which blame or praise attributions, and responsibility attributions more generally, are simply divorced from intentional action attributions, as in the example provided by Mele and Moser (1994, p. 51) according to which a drunk driver may well be held responsible for running over a pedestrian, even though his running over the pedestrian was not intentional. If we take this sort of intentional action (or lack thereof) attribution seriously, we can only interpret Knobe's results as showing that the folk psychological notion of intentional action simply shouldn't be trusted. The view he proposes provides a rationalization of people's attributions of intentional action that only makes sense a posteriori on the basis of subsequent blame or praise evaluations. In line with this idea, Malle and colleagues (Malle, 2006, Malle & Nelson, 2003, reported in Knobe, 2006, p. 217), give the plausible suggestion that people's feelings about blame distort their judgements about intentional actions. Essentially, according to them the order of explanation goes as follows: people judge an agent's behaviour immoral, hence judge her blameworthy, and thus label her behaviour intentional

as a justification to the blame judgement already assigned (Knobe, 2006, pp. 217-218).

After all these considerations, a methodological point is in order. I take Knobe's results to show that if there is any interesting notion of intentional action that matters for philosophical purposes, we are not going to get it from folk psychology, which seems rather to be led astray by moral considerations when it comes to employing the notion of intentional action. I'm therefore going to try to capture a notion of intentional action that, while aiming to be far from a philosophical fiction (see Mele, 2001, quoted in Nadelhoffer, 2005), may still occasionally depart from common sense—for instance, I won't try to rationalize people's attributions of intentional actions as they appear in Knobe's results, but will rather consider them misguided. I think that we're going to get a much more illuminating account of intentional action if we steer clear of considerations to do with blame and praise, and if we rather look for the bases of intentional action attribution in the connection between intentional actions and intention. To this I now turn.

2. First attempt: reducing intention to intentional action

So, how are intentional actions and intentions related? Consider the following reflection:

> how do we tell someone's intentions? or: what kind of true statements about people's intentions can we certainly make, and how do we know that they are true? That is to say, is it possible to find types of statement of the form 'A intends X' which we can say have a great deal of certainty? Well, if

> you want to say at least some true things about a man's intentions, you will have a strong chance of success if you mention what he actually did or is doing. For whatever else he may intend, or whatever may be his intentions in doing what he does, the greater number of the things which you would say straight off a man did or was doing, will be things he intends.
> (Anscombe, 1957, pp. 7-8)

Anscombe suggests that intentional actions and intentions are so close that the latter may be, as it were, *read off* the former: to reach a belief that has a high likelihood to be true about what a man intends, one would do well to consider what the man in question is doing, because, most likely, he's going to intend to do whatever you'd straight off describe him as doing. But this statement shifts the focus on the following issue: just what would you *straight off* describe someone as doing? What does this qualification—*straight off*—refer to?

Suppose you're presented with the following scenario: Mary and John are standing in front of each other, and Mary is pointing to her side, away from them both, while angrily and loudly telling John that he should never show up again in front of her. What is Mary doing at that point? Well, she's telling John that he should never show up again in front of her, she's screaming loudly, she's pointing to her side, and she's also moving a lot of air molecules[31]. Anscombe's appeal to things that we'd straight off say that someone is doing is meant to rule out descriptions such as the latter, according to which Mary is moving a lot of air molecules, insofar as it's going to be "a very recondite piece of information" to most people (Anscombe, 1957, p. 8). The sort of descriptions that, according to Anscombe, matter for the purposes of arriving at what one intends to do are the

[31] This example draws inspiration from Searle (1983, pp. 98-100).

sort of things that we'd be expected to report in a law court on being asked, as witnesses, what a certain person did or was doing (Anscombe, 1957, p. 8).

Admittedly, none of this might sound especially illuminating if we're interested in pinpointing all and only those descriptions of what someone is doing that may be a good guide for us to find out what she intends. Still, at this stage, Anscombe is only interested in, however approximately, getting across the following idea: as observers, we are generally in a good enough position to form judgements as to what a man is doing that are relevant towards his intentions. In the previous example, most people closely observing John and Mary would readily be able to say that Mary is telling John that he should never show up again in front of her, that she's screaming loudly, and that she's pointing to her side. That there are a number of things that we can easily tell about what someone is doing at any given time is, indeed, something that it seems reasonable to accept. If Anscombe's idea, that a man's actions are the privileged way to get to his intentions, could be made to stand, it would follow that we're generally in a good enough position to judge what a man intends just on the basis of what we can easily observe him doing—against the intuitively appealing idea that "what a man's intentions are is only authoritatively settled by him" (Anscombe, 1957, p. 9)[32]. This makes Anscombe's idea an attractive one, and worth considering.

In order to be able to rely on Anscombe's proposal, we first need to clarify the following. Not everything that a man does at any given time is going to be an intentional action of his. Suppose, for instance, that you and I see a man thrashing about in the grip of an epileptic seizure. If you ask me what the man is *doing*, I can

[32] Studies on behaviour parsing (Newtson, 1973; Kurby & Zacks, 2008), hinted at in chapter 6 (section 4) tell us that there's more to the idea that observers are in a good enough position to judge what a man is doing than meets the eye.

tell you precisely that he's thrashing about under an epileptic seizure, and this will be an appropriate answer to your question to the extent that it gives a correct description of this man's behaviour. Strictly speaking, though, I haven't described anything that the man in question is *doing*, insofar as he's not acting intentionally: he's rather best described as there being something *happening to him*[33]. So, if Anscombe's suggestion, according to which a subject's intention has to be worked out from her intentional actions, is to be any use, we're then going to need a criterion for telling intentional from non-intentional actions.

Indeed, Anscombe does try to provide the requested criterion. Her proposal is that an intentional action is one "to which a certain sense of the question 'Why?' is given application" (Anscombe, 1957, pp. 11-ff.). Since we're going to find that Anscombe's idea cannot be made to stand, I'm not going to dwell on her criterion for intentional actions beyond what is actually going to prove relevant for the rest of my enquiry into the notion of intention. In particular, I'd like to focus on a specific way in which, according to Anscombe, the question "Why?" *refuses* application. This is when the person interrogated as to why she's doing something replies that *she wasn't aware* that she was doing it.

Now, the notion of not being aware that one is performing a certain action deserves qualification. Recall the example of Mary telling John that he should never show up again in front of her. Suppose that someone comes along and tells Mary: "Oh, I see you're throwing a real temper tantrum there!" At that point, if Mary

---

[33] The problem of action is to explicate the contrast between what an agent does and what merely happens to him, or between the bodily movements that he makes and those that occur without his making them.
(Frankfurt, 1978, p. 157)

happens to never have heard the word *tantrum* before, she may well reply: "Really? I didn't know I was." As Anscombe (1957, pp. 11-12) points out, it is generally the case that one and the same action can have many different descriptions[34], and someone might know his own action under one description and not under another. In particular, it may well be the case that Mary knows her action as one of *making a scene*, and even as one of *giving John a tell-off*, still without knowing it as one of *throwing a temper tantrum*, simply because she's never heard that expression before. But once she learns that a temper tantrum is no other than an angry outburst, then it's not the case that she was not aware of doing such a thing—she knew it very well, and simply didn't know that specific term to describe it. By contrast, suppose that, by screaming at John, Mary makes it incredibly difficult to concentrate for a student whose window gives onto the street where the scene is taking place. At some point, the student loses his patience, opens the window and angrily asks Mary why she's driving him insane. Now, in this situation she may legitimately answer that she wasn't aware that she was doing such a thing, so that the question refuses application: since Mary didn't mean to drive the student insane in the first place, the question as to why she's doing it cannot even be raised. So, according this piece of reasoning, it follows that being aware that one is performing a certain action is a necessary condition on performing it intentionally[35].

---

[34] The idea that one and the same action may be picked out by different descriptions is not accepted by everyone. Anscombe, for one, subscribes to this thesis, and I limit myself to reporting her view as being in line with it without endorsing this thesis myself. See chapter 6 (section 2.1) for an alternative view on individuating actions (due to Goldman, 1970) according to which, roughly, there is an action type for each different available description.

[35] Anscombe (1957, p. 12) points out that, however, awareness is not sufficient. Take tics an example (1957, p. 13): suppose that someone who suffers from tics suddenly winks

Let us now step back and look at the bigger picture again. We've seen that Anscombe's proposal is that intentions might be somehow read off a subject's intentional actions. If this were true, then we would need just one theory that explains both intentions and actions. That is, a theory of intention could be reduced to a theory of action. As a matter of fact, one of the early versions of the leading theory in philosophy of action, the Causal Theory of Action (Davidson, 1963), held precisely something along these lines. Consider the fact that there are things that we do *with the intention* of doing other things—one of Davidson's examples is going to church *with the intention* to please one's mother (Davidson, 1963, pp. 689-690). According to Davidson, intentions-with-which are simply ways of re-describing what someone is doing in terms of a *primary reason*, conceived as a pro-attitude towards actions having some feature *F*, together with the belief that the action in question has that feature. In the reported example, *pleasing one's mother* would be the feature that the subject believes the action of going to church to have. It is by virtue of a relation to a primary reason that actions count as intentional.

This is an example of a reductive theory of intentions-with-which and of intentional actions to a theory of action featuring primary reasons. But, as it turns out, this attempt to reduction neglects the fact that intentions for the future, namely intentions directed to an action that is going to be performed at some point in the future, include intentions from which no actions ensue. Suppose that, while talking to a friend of mine, I learn about a book that she's recently read whose

---

several times in a row, while being fully aware of this repeated winking. Yet, if asked: "Why did you wink several times in a row?" he may well answer: "It was involuntary", which, according to Anscombe, is another way of refusing application to the question "Why?"

content, as she reports it, sounds extremely interesting to me—so interesting that I right away form the intention to buy it as soon as possible, namely the following day as I do my shopping. As it happens, come the following day I forget to look for the book in question, and fail to execute the corresponding action. Furthermore, from then onwards, I end up never buying the book—my intention never materializes. Davidson (1978) recognized that intentions for the future are sometimes cases of *pure intending*, in which no steps of any kind are taken towards the completion of an action. Cases of pure intending constitute examples of intentions that a theory of action is not in a position to capture. Therefore, the reduction of a theory of intention to a theory of action cannot be performed.

3. Second attempt: intentions are just desires

While in the previous section it was shown that a theory that reduces intentions to intentional action, there are reasons why we might still want to retain the idea that there is a close connection between an intention and a corresponding intentional action. Recall that in chapter 1 (section 2) it was suggested that an action may be interpreted as the attempt to bring about a certain outcome via certain means. In the same chapter, in section 3.1, I pointed out that it is an agent's intention that determines what outcome an action is meant to bring about. But how does an agent's intention determine this outcome? What is the relation between an agent's intention and the outcome that she wants to bring about? Recall the experiment by Meltzoff (1988) reported in chapter 1 (section 2). When the subject bends forward to touch the panel with her forehead, it is plausible to think that she intends to *light up the box*. It's also possible to say that *lighting up the box* is an outcome that she wants to bring about. So, the outcome that the subject wants to bring about is

part of the content of her intention, and so it's possible to say that an intention *represents* the outcome to which an action is directed.

This is something that Searle (1983) proposes in his development of the notion of intention, and we can say more about how an outcome is represented by an intention on the basis of his theory. The assumption that Searle works on is that intentions, just like beliefs and desires, are Intentional states, namely mental states characterized by a directedness upon objects and states of affairs in the world (1983, p. 1). Just as when I believe I always believe that something is the case, and when I desire I always desire something, or that something should be the case, so, whenever I intend, there must always be something that I intend to do. Intentional states can thus be analysed as made up of a psychological mode plus an Intentional or representative content (*ibid*., pp. 5-6). If I believe that it is raining, my Intentional state can be analysed as composed of a psychological mode—belief—plus an Intentional content—*that it is raining*. Some Intentional states have a *direction of fit* (*ibid*., pp. 7-8), namely determine a relation between mind and world that has a specific direction, regulating which between mind and world is to conform to which. For a belief to be true, it must conform to how the world is—it has a mind-to-world direction of fit. For a desire or an intention to be fulfilled, the world has to conform to that desire or intention—they have a world-to-mind direction of fit. Let us notice an important commonality that has just emerged between intentions and desires: both intentions and desires are mental states representing their content with a world-to-mind direction of fit.

For those Intentional states that do have a direction of fit, such as beliefs, desires and intentions, the content of an Intentional state is also the *condition of satisfaction of that state*—e.g., my belief that it is raining will be true if and only if it

is raining. We may thus provisionally settle on an account of intentions as mental states with a world-to-mind direction of fit, insofar as they represent outcomes that the subjects wants to bring about, and of intentional actions as the conditions of satisfaction of intentions.

We could therefore say that, while intentions cannot be reduced to intentional actions, the two are still related in the following way: intentions represent action outcomes that may be brought about by means of intentional actions, and intentional actions are the conditions of satisfaction of intentions. To go back to the example given on the basis of Meltzoff's experiment, the subject's intention is to light up the box: lighting up the box is the outcome that the subject wants to bring about. The subject's intentional action of lighting up the box fulfils her intention to light up the box. On the basis of this, it may seem natural to propose the idea that the relation between an intentional action and the corresponding intention is going to be given by what Bratman (1987) terms the Simple View: one's intentionally *A*-ing implies her having the intention to *A*.

Still, recall that earlier on in this section I pointed out, on the basis of Searle's theory, that mental states representing action outcomes with a mind-to-world direction of fit include not only intentions, but also desires. For instance, suppose that I desire to go to London tomorrow at 10 a.m. I have a mental state representing an outcome—going to London tomorrow at 10 a.m.—with a world-to-mind direction of fit, whose condition of satisfaction is that I actually go to London tomorrow at 10 a.m. If I do go to London tomorrow at 10 a.m., I will thereby satisfy my desire to do so. How does my having an intention to go to London tomorrow at 10 a.m. differ from a desire with the same content? If no difference could be found, then there would be no need to postulate a further kind

of mental state—intentions—in addition to desires. The reasons why intentions are indeed different mental states with respect to desires is provided in the following sections, and is based on Bratman's idea that the main role of intentions is to enable us to effectively make and carry out plans, and, because of this role, intentions are subject to constraints that do not likewise apply to desires. As we're going to see, these constraints are also responsible for the failure of the Simple View.

4. Connecting intentions with actions via the Simple View and the Single Phenomenon View

4.1 Why intentions are not just desires: Bratman on intentions as plans

Consider the following scenario. You hear me saying that I'd really like to spend the whole day relaxing in London tomorrow, and that I'd also really like to finish this chapter tomorrow. Presumably, this statement won't strike you as either especially strange, since it's a typical manifestation of a conflict between desires that most people have to face on an everyday basis, or as inappropriate: you'll think that I may well have conflicting desires—I'll just have to make up my mind before tomorrow. Compare this scenario with one in which, by contrast, you hear me saying that I intend to spend the whole day relaxing in London tomorrow, and that I also intend to finish this chapter tomorrow. I take it that your natural reaction (as well as everyone else's) will be to think, roughly, that things don't work this way, and that if I've stated something like this I must be very confused, at the very least, as to the meaning of the verb *to intend*. The problem, as it's soon going to be made more precise, is that one cannot intend to *A* and, at the same time, intend to *B,* if the bringing about of *B* plainly conflicts with the bringing about

of *A*, on pains of being irrational (see Bratman, 1987, p. 32).

I take it that the two scenarios that I've described above point to features of the folk notions of desire and intention that should be both deeply rooted and uncontroversial enough as to have a bearing on the corresponding technical notions. The reason why it sounds wrong to say that one intends to *A* and, at the same time, one also intends to *B,* if the bringing about of *B* and *A* are incompatible is because, implicit in the notion of intention there is a *commitment* to whatever one has claimed she intends to do—something which, by contrast, is absent in the notion of desire. Bratman's (1984, 1987) theory of intentions respects the difference that I've just highlighted by insisting on a series of ideas. First of all, as human beings we're subject to the need for (a) in advance deliberation over which course of action to take among conflicting available ones, and (b) coordination, both intrapersonal (e.g., I need to harmonize my need to finish a chapter with that to sleep, eat, and do something different from all the above) and interpersonal (e.g., if you need to enter our house, I have to make sure that I'm not sleeping when you knock at the door and are without a key) (Bratman, 1987, pp. 2-3). In short, as human beings we need to be able to effectively plan: our plans for the future must be able to be brought to completion by taking into account both the constraints we're subject to and those concerning any other people on whose contribution our plans rely. Secondly, in line with the example given at the opening of this section, intentions are significantly like plans, the latter conceived as "mental states involving an appropriate sort of commitment to action" (Bratman, 1987, p. 29). Working on these assumptions, Bratman proceeds to spell out a series of features belonging to intentions, which, among other things, contribute to setting them further apart from desires.

Connected to the aforementioned committal character of intentions comes another characteristic, which licenses talk of being *settled* on a certain course of action, namely a certain stability or inertia. Once I have formed a certain intention, I will not normally continue to deliberate whether or not to do whatever that intention is about. Once an intention is formed, deliberation is over, and reconsideration is thereby ruled out—at least, in the absence of new information (Bratman, 1987, p. 16). This is one of the things that makes the statement that tomorrow I intend to go to London and I also intend to finish this chapter contradictory. If I *intend* to go to London tomorrow, it means that I've already taken into consideration alternative things to do tomorrow incompatible with my going to London (such as finishing this chapter), and decided that I'm not going to do them. Bratman (1987, pp. 16-17) points out that this ceases to hold if information that I didn't possess at the time of deliberation suddenly becomes available: if, once I've stopped deliberating and I intend to go to London, I am suddenly told that public transport in London tomorrow will undergo serious disruptions, I might then give up my plan. Deliberation is always ended on the basis of a certain body of information—learning something new might bring one to carry on deliberation.

As a consequence of stability, there comes another characteristic of intentions, namely their being *conduct controllers*[36] (Bratman, 1987, p. 16), which consists in the following fact. My intending to go to London tomorrow at the very least it implies that, if nothing stops me (e.g., if I don't fall ill), I will actually go to London tomorrow. I'm not only inclined to do it, but also committed to doing it.

[36] Notice that Bratman doesn't present the conduct-controlling character of intentions as a consequence of their stability. Still, it seems to me that it's easier to justify the former as following from the latter.

Bratman points out that herein lies another difference between intentions and desires. He notes that both intentions and desires are pro-attitudes, in the sense that they can play a motivational role, i.e. they can move us to act (together with the relevant beliefs). Desires potentially influence our course of action, but, when weighed against conflicting desires, they may be abandoned. Bratman (1987, pp. 15-16) provides the example of one's desire to drink a milk shake at lunchtime which potentially influences what one does at lunchtime, but when lunchtime comes one might end up not even trying to have a milkshake (e.g., because one has weighed the desire of having a milk shake against that of losing weight). Unlike desires, which are mere *potential influencers* of action, intentions are also *conduct-controlling* pro-attitudes, i.e. they involve a special commitment to action.

Another (and more fundamental, for the purposes of what is going to follow) characteristic of intentions is that, once formed, they lead to further intentions by way, e.g., of means-end reasoning (Bratman, 1987, p. 17). If I intend to go to London tomorrow, I may thus form the further intention to take a certain means of transport in that direction. Not only I *may* form such intentions, but I *have* to form such intentions between the end of deliberation and the onset of action. If I intend to go to London tomorrow at 10 a.m., between now and 10 a.m. tomorrow I simply have to have worked out how to go there, or I won't actually be able to go. Intentions have the further feature of intentions is the possibility to feature as elements in larger plans. For instance, I could have the intention to go to London as part of a larger plan to visit the UK.

Now, by insisting on the idea that the notion of plan is central to understanding that of intention, Bratman (1987, p. 31) proceeds to point out some requirements on plans that are meant to ensure the effectiveness of a plan. First,

there are a couple of consistency constraints. A plan should be *internally consistent*: it shouldn't be the case that the bringing about of one element of the plan conflicts with the bringing about of another element of the plan—it should be possible to successfully execute the entire plan. A plan that includes travelling on bus n. 1 between 10 a.m. and 11 a.m. and travelling on bus n. 2 between 10 a.m. and 11 a.m. of the same day is not internally consistent. Furthermore, it should be possible to successfully execute the entire plan working on the assumption that my beliefs are true—that is, my plan should be strongly consistent, *relative to my beliefs*. If I intend to go to London and, in order to do so, I set off northward while knowing that London is down South with respect to where I am, my plan fails to be strongly consistent.

Also, my plans should be *means-end coherent*. This requirement is based on the idea that plans are usually formulated as partial, that is, not specifying from the outset the way in which they're going to be carried out. I might plan to go to London while leaving it unspecified whether I'm going to go there by train, or by bus, and so on. For this plan to be means-end coherent, it has to be the case that I fill it in with a specification as to the means to getting there. Means-end coherence doesn't set any constraint on when the missing details of a given plan should be provided, nor does it require that a plan should be supplemented with the details in question all in one go—in Bratman's (1987, p. 31) words, "it is enough that they be sufficiently filled in before it is, by my lights, too late". Once I form an intention, my intention becomes inserted in a web of other intentions and plans, and on this web bear the requirements of consistency and means-end coherence listed above.

Notice that, in line with the idea that intentions may feature as elements in plans, it can be the case that an intention features as one of the premises in a piece

of practical reasoning. This gives me the chance to reintroduce and expand on an idea that was introduced earlier on in this chapter (section 2). In the course of presenting Anscombe's view on the relation between intentional actions and intentions, I dwelled on her suggestion that, for an action to qualify as intentional, the agent must be aware that he's performing it. Now I'd like to suggest that an analogous constraint holds on intentions—specifically, in order to be said to have an intention, one must be aware of having that intention, and, therefore, must be aware of the content of that intention (chapter 3, sections 3.2-3.3, is going to add a series of provisos to this idea).

John Hawthorne (2004, pp. 29–31) argues for the idea that our instances of practical reasoning should only take as premises propositions we know, by means of the following piece of reasoning. Suppose a person were in possession of a lottery ticket, and is offered a penny for it, and suppose this person reasoned as follows:

(i) the ticket is not a winning one;

(ii) so, if I keep the ticket, I won't gain anything.

(iii) If I sell the ticket, I'm going to earn a penny;

(iv) so I should sell the ticket.

But now, unless the ticket holder has managed somehow to find out that his ticket is not a winning one (maybe by hearing the winning number announced, unbeknownst to the potential purchaser), in which case he'd *know* that the ticket is not a winning one, this piece of reasoning is going to be held unacceptable by anyone that's presented with it. Hawthorne (2004, p. 31) concludes that "[at] a rough first pass […] one ought only to use that which one knows as a premise in one's deliberations."

Since Bratman has shown that one's intentions may feature as premises in pieces of practical reasoning meant to establish the means for accomplishing one's intentions, by adding the requirement identified by Hawthorne it follows that a constraint on intentions is that we must be aware that we have them, and therefore must be aware of their content, and so this content has to be introspectable (chapter 3, sections 3.2-3.3 is going to specify that one only needs to be able to introspect the content of one's intentions *at some point*, and not all the time).

Now, I've focused on Bratman's planning conception of intentions, while saying nothing as to his view on intentional actions, and on their relation to intentions. Given that the notion of intention and some of the constraints which it has to obey are now in place, it is time to broach the topic on the relation between intentions in the planning conception and intentional actions.

4.2 Why the Simple View doesn't work (I): the video-games case

Bratman (1987, p. 111) points out that it is part of our ordinary conception that intentionally doing something and having an intention must be connected in some way (and, indeed, sections 2 and 3 of the present chapter have explored this idea), and yet it's hard to specify how. Now, an intuitively plausible idea on what this connection could consist in is the following. While I'm acting intentionally—say, as I'm going to London—there's surely something that I intend to do. Given that what I'm doing is going to London, among the things that I intend while performing that action there will be *going to London* (Bratman, 1987, p. 112). This suggests that the connection between an intentional action and an intention could

be the following: for me to intentionally *A*, I must intend to *A*. This is what Bratman (1987, p. 112) terms the *Simple View*[37]. And here is why it doesn't work.

Bratman (1987, pp. 113-116) builds up his case against the Simple View by means of three steps. In the first step, Bratman supposes that a subject plays a video game in which he has to guide a missile to a target. Bratman supposes that the subject (a) is skilled at playing this video game, and (b) is however doubtful as to his success due to the difficulty of the game. In spite of his doubts, the subject hits the target, in a way that depends on his skills, which are present according to (a), and therefore in a way that is not a matter of luck. It can be thus said that the subject hits the target *intentionally*. This is what the subject has tried to do, and, Bratman underlines, hitting the target depends on the subject's skills at playing this game. On the basis of the Simple View, the subject's intentionally hitting the target implies that the subject *intends* to hit the target.

Now onto the second step. This features the addition of a second video game, just like the one described in the first step. In this case, the subject, who is supposed to be ambidextrous, simultaneously plays two video games involving guiding a missile to a certain target, one with each hand. Once again, (a) and (b) hold: the subject is skilled at playing the video game while being doubtful of success. This time, the subject hits target 1 but misses target 2. As his success with target 1 depends on what he was trying to do as well as on the relevant skills, it can

[37] Bratman (1987, p. 112) hastens to add that the Simple View, while requiring that, at the time at which one is intentionally *A*ing, her mental states must include an intention to *A*, doesn't say that intending to *A* should constitute a separate event with respect to intentionally *A*ing.

once more be said that the subject hits target 1 intentionally. Assuming that the Simple View is correct, it follows that he intends to hit target 1. Bratman points out that since an analogous attempt is being carried out towards target 2, with an equal amount of skill involved, if the subject intends to hit target 1 he must therefore also be said to intend to hit target 2, so that we should conclude that the subject intends to hit both targets.

Trying to follow this piece of reasoning, one may realize that the Simple View leaves open the following: from what point in time does the subject have the intention to do whatever he's going to do intentionally? If this is left open, a strong version of the Simple View could in principle hold, according to which a subject's intentionally A-ing from x_1 onwards implies his having an intention to A from x_1 onwards, and not beforehand. If this version of the Simple View held, we could question the step from (3) to (4), on the basis of the fact that the Simple View doesn't grant that an intentional action beginning at time x_1 implies that the intention to so act could be held during a time interval running from x_0 to x_1, where x_0 strictly precedes x_1. But even assuming a strong version of the Simple View along these lines is enough for Bratman's purposes. Suppose, according to a stronger version of the Simple View, that a subject only has an intention to hit the target only beginning at the time at which the target is actually hit. Note that a possible outcome of the game is that the subject hits both targets. Given the subject's effort and skill, he can be said to hit both targets intentionally, and hence, however momentarily, has the intention to hit target 1 as well as target 2. This is enough for setting up the stage for the third step.

Bratman further complicates the scenario proposed at the second step by means of the following assumption: the two video games are so linked that one

may hit target 1 and miss target 2, or vice versa, or miss both targets, but if both targets are about to be both hit the whole system shuts down, thus making it impossible to hit both targets. It is possible to see both targets until the end of the game, so one can know which one he's hit, if any. Furthermore, there is a reward for hitting either target. The subject, in addition to being skilled at the video game as well as at playing both simultaneously, knows how the two are linked, and also about the reward. Given all this, the subject decides to play both games, in order to increase his chances of getting a reward, and for the same reason tries to hit both targets. By the same reasoning made at the second step, whatever target is hit—say, target 1—can be said to be hit intentionally, and thus the subject can be said to have an intention to hit that target. Again, Bratman says that, because of the symmetry of the case, the subject must also intend to hit target 2. On the basis of the two possible interpretations of the Simple View that I gave at the second step, this is a justified conclusion. But in this case, the subject ends up having two intentions that are inconsistent with his belief that he cannot hit both targets. This is a violation of the strong consistency requirement holding for intentions, which makes the subject irrational. But, by Bratman's lights, trying to hit both target is a perfectly reasonable strategy, so the subject shouldn't be judged irrational. Since what generates the charge of irrationality is the attribution of the two intentions, this attribution is wrong. Since what leads to this attribution is the Simple View, Bratman concludes that the Simple View must be false: I can intentionally A while not intending to A. Bratman takes this example to show that intentions are subject to a web of norms, one including the strong consistency requirement, that does not likewise apply to intentional actions.

4.3 Why the Simple View doesn't work (II): webs of norms

Specifically, Bratman contends (1987, pp. 133-ff.) that typical cases in which I intentionally act—say, I intentionally A—are characterized by the fact that the following three are true:

(a) I intend to A

(b) I endeavour to A

(c) I intentionally A.

These may occasionally come apart insofar as they're subject to different webs of norms. In particular, intentions are subject to demands for *rational agglomerativity*—that is, on the basis of the role played by intentions in terms of plans, it should be possible for intentions to be put together into a larger intention, and it would be rational for an agent to try to put them together in this way. Not so for endeavouring, it turns out. Indeed, the video-games case is an illustration of the fact that it might be rational for an agent to endeavour to hit target 1, and to endeavour to hit target 2, while at the same time it's rational for the agent *not* to endeavour to hit them both. Lack of demands for rational agglomerativity extends from endeavouring to acting intentionally. This, again, can be seen in the video-games case: one may end up hitting target 1 intentionally, as well as hitting target 2 intentionally, thus, unfortunately, hitting both targets, while still not hitting both targets intentionally. The video-games case, in short, provides an example of a case in which intending to A on the one hand and endeavouring to A and A-ing intentionally on the other hand come apart, insofar as the former, but not the latter, is subject to demands for consistency (see section 4.1 of the present chapter) and agglomeration.

Bratman contends that endeavouring is closer to intending than it is to intentional action in some respects. These have to do with what he calls *belief-extendability*. He resorts to the following example (1987, pp. 123-ff.): suppose I intend to run a race, and I believe that, in so doing, I will wear down my sneakers. This is an unwanted, as much as it is unavoidable, consequence of my running the race. Suppose I not only believe that I'm going to wear down my sneakers, but I also notice that I'm doing so as I run the race. On the basis of these two assumptions, Bratman claims that wearing down my sneakers is something that I do intentionally (something that I'm going to challenge shortly). Still (and I agree with this), it would be wrong to say that I intend to wear down my sneakers—witness to this is the fact that I'm not ready to engage in means-end reasoning designed to achieve this result. That is, (Bratman, 1987, p. 135), my belief that I'm going to wear down my sneakers doesn't extend my intention to run the race to an intention to wear down my sneakers. Likewise, says Bratman (1987, p. 135), I do not endeavour to run down my sneakers: my belief that I'm so doing doesn't extend my endeavour to run the race to an endeavour to run down my sneakers.

Now I'm going to challenge the idea that the action of wearing down my sneakers in the previous example should be classified as intentional. Mele and Moser (1994, p. 45) point to an example due to Gilbert Harman (1976, p. 433), according to which a sniper, in the attempt to kill a soldier, knowingly acts in such a way as to render an enemy aware about his presence. Mele and Moser rightly make the following considerations. Since the sniper is aware that he's going to alert the enemy to his presence—he doesn't do it by mistake or by accident—we cannot say that he *unintentionally* alerts the sniper to his presence. Still, our description would somehow be incorrect if it said that the sniper *intentionally*

renders the enemy aware about his presence, for the sniper does not aim at rendering the enemy aware about his presence in any way. Cases consisting of known and unwanted consequences of one's own action are captured by Mele and Moser in terms of a middle ground which they term *nonintentional* actions, which does justice to the fact that one does not aim to that consequence at all, and doesn't take any steps in order to foster its realization (if anything, one will take steps to prevent that consequence). The case of wearing down one's sneakers qualifies as more aptly located in this middle ground.

Even if we might not (as I am not) ready to accept the attribution of intentional action in the sneakers case, and, if so, we might want to challenge the need to separate out the notion of endeavouring from that of acting intentionally, on the basis of the missile case Bratman has still enough grounds for rejecting the Simple View, on the basis of the idea that intentions are plans and so are subject to a different web of norms with respect to intentional actions and endeavouring (Bratman, 1987, §§8-9). Bratman, however, still allows that there is a significant connection between intentionally *A*ing and some intention, though not necessarily the intention to *A*. He thus proposes what he terms the Single Phenomenon View (1987, pp. 119-ff.), such that, whenever one intentionally *A*s, one's intentional action is within the *motivational potential* of an intention to *B*, where *B* isn't necessarily identical with *A*. As for a non-trivial case in which *A* differs from *B*, the video-games example is an illustration of the possibility of intentionally hitting (say) target 1 as part of the motivational potential to carry out my intention, which is *not* straightforwardly to hit target 1[38]. This view avoids difficulties such as that introduced via the video-games case, while retaining the idea that an action *A* is

[38] As to a positive characterization of what I intend in the video-games example, Bratman (1987) considers different options at pp. 120-122.

intentional by virtue of being suitably related to some intention, though not necessarily the intention to *A*.

6. "What good is a will?"

Taking stock, we might settle on a provisional view of intentions that sees them as mental states that have the following characteristics:

(i) they represent action outcomes with a world-to-mind direction of fit (section 3; chapter 1, section 3.1; Searle, 1983);

(ii) they enable effective planning, and are inserted in a specific web of norms that differentiates them from other propositional attitudes (sections 4.2-4.3; Bratman, 1987)—specifically desires, which have the same direction of fit;

(iii) endowed with a content that is accessible to introspection (by virtue of their role in planning; Hawthorne, 2004, pp. 29-31).

What we can conclude on the basis of the previous sections, and particularly of Bratman's theory, is that intentions play a motivating and a conduct-controlling role in action production—a role that seems to be especially salient when we are still at the planning stage of a certain action, or, as it were, when our intentions are still *future directed*. To put it in the way in which Velleman (2007, p. 195) sums up Bratman's theory, we can say the following things about intentions:

(1) future directed plans are the paradigm of all intentions;

(2) intentions enable one to organize deliberative effort by making it possible to deliberate in advance, when there are enough resources (in terms of time and information) to do so, and to preserve the results of our deliberation until the time for acting comes, when deliberation might prove more difficult;

(3) intentions make action coordination easier, both for the agent herself (at the intrapersonal level) and in relation to other people with whom the agent might wish to coordinate (at the interpersonal level).

Bratman also subscribes to the Single Phenomenon View, according to which:

(4) whenever one intentionally *A*s, one has an intention—the intention to *B*, where *A* and *B* are not necessarily the same.

It is certainly true that, on many occasions, our actions are preceded by a, however brief, phase of making up our mind. On those occasions, it is plausible to think that our actions are informed by a kind of mental states—intentions—endowed with the characteristics described by (1)-(3). It is therefore also plausible to think that (4) holds on those occasions. Still, consider that there are situations in which no deliberation at all seems to take place. Velleman (2007, pp. 197-198) raises the following doubt: what room is there for intentions conceived according to (1)-(3) in those situations? First of all, a distinction needs to be made among different cases of actions seemingly not preceded by any deliberation. One is given by sudden instinctive actions, such as reaching up and catching a ball thrown at me without warning (Bratman, 1987, p. 126; reported by Velleman, 2007, p. 197).

Bratman claims that catching a ball in these circumstances (i.e., in which one is instinctively reacting to an event that catches her us by surprise) fails to qualify as an intentional action, insofar as, although it is under the subject's control and is not a reflex, it is however too "automatic and unreflective" (Bratman, 1987, p. 126) to count as intentional, though it's not unintentional. While more would probably need to be said to be justify this interpretation of spontaneous actions, let us accept it for the sake of argument[39].

There is another class of situations that do not seem to be preceded by any deliberation are exemplified by the following. Suppose that we're at a party sitting on a chair and talking to someone, when suddenly we're offered a plate of cookies.

> The question is why we bother to form intentions when there is no longer any opportunity for them to serve the functions for which, according to Bratman, the mental state of intention is designed. When the plate of cookies is held out to us, why do we make up our minds to take one? Why doesn't our hand just shoot out and grab, as it does when we spontaneously and automatically react to a sudden throw?
>
> (Velleman, 2007, p. 197)

This worry is based on the hypothesis that the functions of an intention may be exhausted by (1)-(3). If that hypothesis were true, it would be unclear what function an intention could play once the time for acting comes, at least for actions

[39] Note that Bratman (1987, p. 126) suggests also an alternative account of this kind of actions, an account that characterizes them as the execution of a long-standing policy—in the case of the catching the ball example, e.g., that of protecting oneself. According to this interpretation, these actions would still involve an intention, although one with a general content, not tailored to the specific case. It is not clear that this interpretation would fare any better in countering the objection that is going to be raised shortly.

such as grabbing a cookie as in the example by Velleman (2007). I'm going to suggest (along the lines of what has been done by Searle, 1983, and by Pacherie, 2000, 2003, 2006, 2008) that if we think that the functions of intentions are exhausted by (1)-(3), we'll only see half of the picture as far as the connection between intentions and actions is concerned. In what follows, I'm going to give hints as to what the other half of the picture is going to be about. A proper exploration of this other half is going to occupy the rest of the thesis, and will help me answering the question as to how actions are connected with intentions.

Causal theories of action, originated with Davidson (1963), have it that a given piece of behaviour counts as an action by virtue of its prior causal history. This may imply, and indeed has been taken to imply (e.g., by Frankfurt, 1978) that actions and mere bodily movements or happenings do not essentially differ at the time of their execution[40]. Frankfurt thinks that this is a mistake, insofar as, by focusing on the causal antecedents of actions, causal theories ignore and do not account for

> the most salient differentiating characteristic of action: during the time a person is performing an action he is necessarily *in touch with the movements of his body in a certain way*, whereas he is necessarily *not in touch with them in that way* when movements of his body are occurring without his making them.
>
> (Frankfurt, 1978, p. 158, my emphasis)

[40] This, according to Frankfurt, is what gives rise to the so-called *wayward causal chains* (see Davidson, 1973, pp. 153-154, reported in Searle, 1983, p. 83), which I'm going to mention later on in this section.

The difference between a persons's being and not being "in touch with the movements of his body in a certain way", or between movements being or not being "*under the person's guidance*" (1978, p. 158) certainly picks out distinct real phenomena, which can be illustrated as follows. Frankfurt makes reference to the thrashings about of a person's body during an epileptic seizure (1978, pp. 158-159). In the latter case, clearly, the person is *not* controlling the movements of his body. Compare a man's raising his arm with his arm going up as a result of the stimulation of his motor cortex. It could be the case that he is strongly opposed to his arm going up, but still can't prevent it from moving once his motor cortex is stimulated. Again, the man is *not* controlling the movements of his body.

Could intentions have anything to do with the aforementioned expounded difference between bodily movements being, or failing to be, under a person's guidance? Notice that there is a tight connection between this question and that as to whether an action is intentional or not. Recall Anscombe's characterization of intentional actions as those to which a certain sense of the question "Why?" is given application. This question may well be applied to those bodily movements that are under a person's guidance, and those are the only ones to which it can be applied: by contrast, movements that are not under the person's guidance, such as an epileptic's thrashing about, cannot be accounted for in terms of reasons why one made them—there is a sense in which the person didn't make them at all.

We have so far explored the *motivational* contribution of intentions to actions—we perform certain actions because we want to fulfil certain intentions—but it would be very strange if an intention didn't have anything to *causally* contribute to the execution of the corresponding action. Suppose that there is a complex of neurophysiological mechanisms that are responsible for the execution

of a given action A—call this complex M_A. Now call I_A the intention to A. It's hard to imagine that M_A and I_A could bring about the execution of A in complete mutual independence. If they did, there would be the danger that M_A and I_A could pull the relevant agent in different directions. So there must be some kind of connection between them[41]. So, there are reasons for thinking that there must be a causal connection between an intention and the corresponding action. But how should the causal link between intentions and actions be characterized? Is it to be conceived along the lines of a billiard ball collision? That is, do intentions simply initiate the corresponding actions, so that their causal contribution ceases with the onset of action? Or does their causal role extend beyond action onset?

Searle (1983) has provided a causal account that features two kinds of intentions: prior intentions, and intentions in action. Prior intentions are characterized as intentions that are formed prior to the initiation of an action, whereas intentions in action are conceived as causes of the bodily movements by means of which a certain action is executed. Not all actions have prior intentions, whereas all actions have intentions in action. The causal picture goes as follows: a prior intention (where there is one) causes an action, where the latter is made up of an intention in action causing a bodily movement. So prior intentions indirectly cause the bodily movements constituting a certain action, through an intention in action.

Some examples of so-called *wayward causal chains* illustrate why an event may fail to qualify as an intentional action in spite of the presence of the relevant prior intention. A climber who wants to get rid of the weight and danger of holding

[41] A similar piece of reasoning is put forward by Butterfill and Sinigaglia (2012) to the effect that an intention and a corresponding motor representation must be related to each other in the context of a subject's bringing about a given action.

another man on a rope (Davidson, 1973, pp. 153-154) might be so unnerved by these reflections as to end up loosening his hold. But loosening his hold, in this case, is not under his guidance. It just happens to him. Searle's interpretation is that the climber doesn't have the relevant intention in action. This lack of being under the guidance of an intention in action is what makes the event of the climber loosening his hold fall short of being an intentional action. As noted by Pacherie (ms.), an intention in action could be considered precisely this: "a placeholder for whatever it is that makes an action intentional." On the basis of these considerations, it should be plausible to think that intentions might do some important explanatory work in the characterization of intentional action. More to the point, it is plausible to think that the explanatory work of intentions is based on a function that intentions fulfil during action execution—which would enable us to answer the question "What good is a will?" even for spontaneous actions that do not seemingly involve any prior deliberation. For this potential to be actualized, though, more needs to be said about what Searle terms *intentions in action*, going well beyond Searle's characterization and, as Pacherie (2000, 2003, 2006, 2008) suggests, helping ourselves to the cognitive neuroscience of action whenever that can help us.

§3. How are intentions and motor representations related?

Introduction

Searle's (1983) dual theory of intention distinguishes between prior intentions, which may be assimilated to intentions as they are standardly conceived (e.g., by Bratman, 1987), and intentions in action. He characterizes intentions in action as causes of bodily movements (1983, p. 95), namely as something that "proximately causes the physiological chain leading to overt behaviour" (Pacherie, 2000, p. 403). Work in the neurophysiology of action provides us with the notion of *motor representation* (see Jeannerod 1994, 2006), consisting in the representation of an action outcome that specifies the pattern of movements that the subject is going to perform, and which thereby drives action execution. More will be said in the course of this chapter (section 1) about the characteristics of motor representations, and in particular it will be shown that, insofar as motor representations play a key role in the production of behaviour, they are obvious candidates for fulfilling the role of intentions in action—a position endorsed by Pacherie (2000, but see also 2003, 2006, 2008) and motivated thus:

> The reason for equating Searle's intentions in action with Jeannerod's motor representations is that they are assigned the same function in both models, i.e. they are the proximal causes of actions and they play a continuing causal role in shaping the action, guiding and monitoring it until completion.
>
> (Pacherie, 2000, p. 409)

On these assumptions, should we accept the conclusion that motor representations are intentions?

Some would reject this idea. Bach (1978), for instance, suggests that what he terms *effective representations*, the characterization of which is rather close to that of motor representations, may underlie the execution of voluntary movements. Still, Bach denies that effective representations are intentions:

> Effective representations are not intentions. Aside from characteristically not being conscious, unlike intentions, they need not represent the behavior as an action on the part of the agent. Moreover, effective representations are more fine-grained than intentions, representing not the behavior as a whole but merely the next bit of behavior (insofar as we can legitimately speak of bits).
>
> (Bach, 1978, p. 367)

Butterfill and Sinigaglia (2012), on the other hand, argue that, while some motor representations are relevantly similar to intentions insofar as both represent action outcomes, still, motor representations are also unlike intentions insofar as they have a motor format, which makes it impossible for motor representations to be inferentially integrated with intentions. Since intentions should be amenable to mutual inferential integration, motor representations cannot be intentions. In short, arguments can be found to the effect that motor representations bear too distant a relation to intentions standardly conceived (e.g., Bratman, 1987) to deserve that qualification.

So, there are reasons, on the one hand, to maintain that motor representations are intentions, and, on the other hand, reasons to think that motor

representations are the wrong thing to be identified with intentions, although the two must be somehow related. In this chapter I'm going to argue for an intermediate, neutral position according to which, although intentions and motor representations are in principle distinct, it is however the case that the content of some motor representations can be a good guide to the content of some intentions, which, following Jeannerod (1994) and in a different sense with respect to Pacherie's (2003, 2006, 2008) use of the term, I'm going to call *motor intentions*. My stance leaves open the possibility that some motor representations could be intentions, but none of what follows is going to hinge on this possibility.

Pacherie proposes the distinction between two kinds of intention whose function is to initiate, sustain, guide and monitor actions: present-directed intentions (or P-intentions) on the one hand, and motor intentions (or M-intentions) on the other hand. These two kinds of intentions are supposed to differ, among other things, in terms of possible contents—the contents of P-intentions being accessible to introspection, while the contents of M-intentions are typically not. I partially agree with Pacherie as far as the possible contents of intentions are concerned. More specifically, I endorse the idea that the content of intentions can encompass all the contents that she describes as pertaining to P-intentions and at least some of the contents that she describes as belonging to M-intentions, with the exclusion of contents concerning action adjustments. Still, I reject the notion that two kinds of intentions should be distinguished on the basis of accessibility to introspection.

1. What are motor representations and why think they could be intentions?

In what follows, the notion of motor representation will be presented as the encoding in the brain of an action, or rather of some of its parameters, prior to its execution. The message that this section is meant to convey is that there exists evidence for this kind of in advance encoding, regardless of the amount of detail in which an action may be represented in advance, which deserves careful but separate treatment. I'm going to stress the amount of independence that action encoding seems to enjoy with respect to the sensory input that may lead to movement corrections, a point that will become relevant towards the end of this chapter.

The concept of motor representation (see Jeannerod, 2006, pp. 8-ff. for a review) was introduced in motor physiology at the end of the nineteenth century, in response to the sensory-motor theory of action generation. According to the latter theory, actions would in one way or anther consist in reactions to changes in the environment. While this view is reasonable as far as reflex movements are concerned, it is not straightforwardly applicable to the notion of a voluntary movement, insofar as it's not obvious to what extent, if any, the generation of a voluntary movement should depend on sensory input. What lent plausibility to the sensory-motor theory of action generation was a series of deafferentation experiments carried out on monkeys (Mott & Sherrington, 1885). These experiments involved the suppression of sensory input from a given limb to the central system, by means of a section of the dorsal spinal roots on the side corresponding to that limb. As a result, the deafferented limb was almost paralyzed, and could only be used to produce awkward movements if the monkey was forced to resort to that limb. Based on these observations, Mott and

Sherrington's conclusion was that movement initiation and execution depend on sensory input to a large and significant extent.

Revision to the sensory-motor theory of action generation was subsequently suggested, and Lashley was among the main proponents of this revision on the basis of the following observations. A patient with a deafferented leg, who thus received no sensations from that leg, was able to bend his knee at a given angle, or to bring his foot at a given height that the experimenter had indicated (Lashley, 1917). The absence of sensations from the deafferented leg excludes that the patient could be relying on sensory input from that leg for the production of the requested movement. The patient was also blindfolded, something which rules out a contribution of visual feedback to the successful performance of the given task. A compelling interpretation of these data suggests that there has to be an encoding of the action in the brain that is going to enable action performance independently of sensory input—this is what I'm going to refer to as a *motor representation*. Bizzi and colleagues (Bizzi et al., 1971) carried out more deafferentation experiments in support of the notion of motor representation. They showed that a monkey whose forelimbs were deafferented was still able to perform elbow movements directed to a visual target, while not seeing the limb, with reasonable accuracy.

Lashley (1951) found further evidence in support of the notion of motor representation in the rapidity of many movements performed by typically developed subjects. When one plays a musical instrument, it is possible to alternate one's fingers reaching the frequency of 16 strokes/s, something that makes the influence of any sensory feedback on the production of those movements impossible. Hence, Lashley concluded that the succession of those

movements had to be encoded prior to action execution. Note that another (related) conclusion follows: not only the movements have to be encoded in advance, but they are also seemingly insensitive to sensory feedback, so that what we call *motor representation* in this case does not encompass movement corrections that might be induced by this feedback.

Thus far, evidence has been brought in support of the idea that motor representations exist, which encode some features of the corresponding action prior to its execution. Within the context of some of the evidence reported, it was hinted that these motor representations seem to be independent from the kind of sensory feedback that might prompt movement corrections. Now notice that two things haven't been established yet. First, it hasn't been explained how actions are generated on the basis of motor representations. This is a problem that would take us too far afield, and that will therefore be left aside (possible solutions to this problem are proposed, e.g., by Wolpert et al., 1995 and Wolpert and Ghahramani, 2000). Secondly, I haven't addressed the question as to what the content of these motor representations is—that is, what features of the corresponding action they specify, and in what detail. An answer to this question, by contrast, is crucial for my purposes, and will only be partially addressed in the following sections, while a more comprehensive treatment of the content of motor representations will have to wait until chapter 4. The provisional take-home message is the following: motor representations exist, and their defining characteristic is the representation of at least some features of the corresponding action prior to the execution of the latter. To the extent to which motor representations represent action outcomes, they have something essential in common with intentions (see Butterfill & Sinigaglia, 2012). Can we draw the conclusions that they *are* intentions, then?

2. Are motor representations intentions?

In the previous section it was established that intentions and motor representations are alike in that they both represent action outcomes. The question then naturally arises, then, as to whether motor representations are intentions. In this section, I'm going to show that intentions may influence those motor representations that represent action outcomes, and vice versa. This double influence, while not constituting sufficient grounds for claiming that motor representations are intentions, still points to a close connection between the two. This connection, as it'll become clear in the course of this section, often seems to be provided by motor imagery, namely "the ability to generate a conscious image of the acting self" (Jeannerod, 2006, p. 24). This ability, which seems to mediate the reciprocal influence between intentions and motor representations, provides the grounds for settling on the view that intentions and motor representations, regardless of whether they should be identified with each other, exhibit an interesting commonality in terms of content—which is going to warrant our talk of, e.g., *the intention to grasp*, as well as provide the foundations for an answer to the question as to how intentions connect with actions.

Before I proceed to show the reciprocal influences between intentions and motor representations, a few words on motor imagery are in order. Motor images (Jeannerod, 1994; 2006, pp. 23-ff.) are mental images of the unfolding of an action that the subject imagines herself executing. As such, they involve a conscious content, which they share with mental images in general, as well as an unconscious content that preserves many of the features of executed actions. Early evidence in favour of motor images preserving features of actions therein represented was

provided by Landauer (1962), who showed that the physical and the mental performance of the same action took approximately the same time (his observations concerned the action of reciting a series of numbers). The fundamental feature of motor images, for the purposes of my argument, is that they are significantly similar both to intentions and to motor representations. As for the former, the content of motor images is at least partially accessible to introspection. As for the latter, motor images display striking commonalities (in terms of timing, encoding of programming rules and biomechanical constraint, as well as in terms of shared neural resources) with the actions that are imagined, which leads us to think that an enquiry into the content of motor images is going to shed light on the content of motor representations.

As for the extent to which intentions influence some motor representations, I'm going to consider evidence based on the mechanisms underlying everyday use of objects for different purposes, and on the disruptions of these mechanisms (specifically, on the utilization behaviour syndrome). To begin with, note that an object, such as a fork, can be used for different purposes—e.g., for eating, or for forcing a drawer open—and, on the basis of it, grasped in different ways.

> The act of grasping a fork, [...] (see Milner and Goodale 1995, 203), requires not simply the provision of an accurate precision grip, but a grip appropriate to the intended use of the fork.
> (Clark, 2001, p. 510)

Now consider the following piece of evidence (reported in Pacherie, 2006, p. 152). When one views an object, even in the absence of any intention to act, the possibilities for action afforded by that object are automatically detected, and they

prepotentiate the corresponding motor programs (Tucker & Ellis, 1998; Grèzes & Decety, 2002), where motor programs may be identified with motor representations. The motor program that is most suitable for the intended use may be selected in different ways. One possibility, reported by Pacherie (2006, p. 152), is proposed by Shallice (1988): different motor programs compete and the one that shows the strongest activation is triggered as a result of a process called *content scheduling*. Another possible way of selecting the most appropriate motor program would consist in explicitly reflecting on the different ways in which one could act upon, and therefore grasp, a given object—something that might happen, for instance, in an unfamiliar situation. As a rather extreme example for this kind of situation, suppose that I intend to grasp a heavy vase that's very ornate and very fragile to move it somewhere else, making sure I don't break it. I might spend some time considering in which point(s) I should grasp that object, with what kind of grip, and so on. Thus and in other ways intentions may lead to the selection of appropriate motor representations.

Conversely, a disruption at the level of intention formation may lead to a disruption in the activation of motor representations. This seems to be the case with patients exhibiting *utilization behaviour* (Lhermitte 1983; reported in Jeannerod, 1994, p. 200). These patients present prefrontal lesions that may thought to be responsible for an impairment in inhibitory control, which may ultimately be interpreted as an impairment of intention formation. This leads to the compulsive imitation of gestures or of complex behaviour that unfolds in front of these patients, or to their compulsively using common objects placed in front of them. For instance, a patient exhibiting utilization behaviour, when presented with a glass and a bottle of water, would continuously pour water in the glass and drink

from it. In the light of the fact mentioned earlier on that the presence of affordances may lead to the automatic prepotentiating of corresponding motor programs (see Pacherie, 2008, p. 186), utilization behaviour may be interpreted as a disruption of intentions, manifested in a lack of inhibitory control, which leads to a disruption of motor representations, consisting in the compulsive recruitment of these representations. Interestingly as for the involvement of motor imagery in the mediation between intentions and motor representations, Jeannerod (1994, p. 200) points out that it is possible "that frontal patients with this syndrome should be unable to generate motor imagery without immediately transferring the imagined action into motor output."

Now I'll show that the reverse kind of influence, namely from motor representations representing outcomes to intentions, exists. I'm going to resort to evidence concerning apraxic patients (reported in Jeannerod, 1994, p. 200; 2006, pp. 12-ff., and in Grafton & Hamilton, 2007, pp. 596-ff.), which, as will soon become clear, can be interpreted as witnessing alterations in the patients' motor representations, though at a level higher than that of elementary motor representations[42]. Liepmann coined the term *apraxia* to indicate a disorder characterized by the failure to execute complex actions, namely actions involving the use and organization of more elementary motor representations in a sequence. Liepmann (1905, reported in Jeannerod, 2006, p. 12) assumed that these patients had lost what he termed *movement formulas* (a notion relevantly similar to that of motor representation). Following Liepmann, Jeannerod (2006, p. 12) defines apraxia "as the consequence of a disruption of the normal mechanisms for action representations."

[42] Cf. an alternative interpretation of apraxia put forward by Pacherie (2011, p. 71).

This disorder shows up at a level that is higher than that of elementary motor representations. The ability to perform elementary actions is preserved—for instance, apraxic patients have no difficulties when it comes to executing simple actions such as grasping an object (Jeannerod, 2006, p. 12). Their difficulty seems to reside at a higher level, that of action organization. As Liepmann emphasized (see Grafton & Hamilton, 2007, p. 596), apraxia is not so much a disruption of movement, but rather of purposeful behaviour. To explain what the difference between movement and purposeful behaviour amounts to in this context, and to characterize what the level of action organization consists in, recall the earlier reflection about the fact that an object may be grasped in different ways depending on the use one intends to make of it. Apraxic patients fail precisely in the selection of the motor representation that's most appropriate to the use of a given object. Clark et al. (2004, reported in Jeannerod, 2006, p. 13) found that apraxic patients fail to pantomime correctly the action of slicing bread in the absence of both bread and knife, due to the incorrect orientation of their movements and to the deficient spatiotemporal coordination of their joints. Should you think that the problem with these patients lies in the inability to imagine the action without the tool that they're meant to use, which would still be a significant finding, and point to the role of motor imagery in action planning, Ochipa et al. (1997, reported in Jeannerod, 2006, p. 13) collected the following evidence on patient G.W.: not only did she fail in pantomiming the use of 15 commonly used tools following just a verbal instruction, but she also failed when asked to imitate an actor, when shown the object but prevented from using it, and even actually handling the object in question. Yet G.W. could correctly distinguish objects according to their function. The data concerning G.W. are especially interesting for

our purposes in that she's not only impaired in action execution, but also in the generation of the motor imagery related to the execution of a given action. That is, if asked what posture her hand would take when performing an action, she was unable to answer.

The difficulty of apraxic patients at the level of action organization is also manifested in their inability to plan actions that require a number of steps to be performed. Lehmul and Poeck (1981, reported in Jeannerod, 1994, p. 200) observed that apraxic patients cannot put cards portraying different stages of a common complex action such as preparing tea in the correct order. This is a demonstration of the inability to organize elementary motor representations in a sequence as required by the planning and execution of a complex action.

Furthermore, experimental evidence concerning aplasic individuals (Gazzola et al., 2007, presented in chapter 1, section 3.2.5) can be interpreted as showing an influence of motor representations on intentions—specifically, in this case, on intention recognition. The main finding of this study is that both aplasic individuals, while observing hand actions, activated those areas in their brain—part of the mirror neuron system—that had previously been shown to be involved in the execution of foot or mouth actions. A plausible interpretation of this finding would have it that aplasics recognize typically developed individuals' intention to execute hand actions in terms of the *motor representations* that they, as aplasics, would recruit to execute the same actions with the effectors that are actually available to them, namely mouth and feet. Their mouth and feet motor representations, that is, influence the recognition of typically developed individuals' intentions to execute hand actions.

The presence of reciprocal influences between some intentions and some motor representations, as shown in the course of this section, points to a close connection between the two, while not warranting the conclusion that the two should be identified, and thus helps us to shed light on how intentions and actions are connected. Still, some of the data reported suggest an involvement of motor imagery in the mediation between intentions and motor representations, whereas those data and examples for the explanation of which motor imagery was not explicitly invoked might still be interpreted as involving motor imagery. Consider again, for instance, the example concerning the lifting of a heavy, ornate and fragile vase. One may try out different possible grips before proceeding to lift it, or simply mentally rehearse those different kinds of grips. Therefore, while I'm going to remain neutral on the question as to whether motor representations are intentions, I've given reasons for thinking that an interesting overlap in terms of content, which could be provided by motor imagery, seems to hold between the two. Just how much overlap there is, and what consequences this has for views on intentions that take into account motor representations such as Pacherie's (2000, 2003, 2006, 2008) will be tackled in the course of the following sections.

3. What's the relation between the content of intentions and the content of motor representations?

Suppose you're about to pick a piece of fruit from a basket – you let your hand wonder over the basket for a bit, then choose – say – a plum, and finally grasp it. This is something that one normally does rather effortlessly, and hardly paying attention to what he's doing. You may well be absorbed in a conversation with someone else sitting at the same table and yet successfully grasp the selected piece

of fruit. Suppose I now tell you that, whether you're aware of this or not, whenever you're about to grasp something, your fingers stretch pre-shaping in accordance with the selected object. They do so in such a way that the grip size quickly reaches a maximum, which is proportional to the shape of the anticipated size of the object, before fingers flex again to match the size of the object (Jeannerod, 1986, p. 104). Presumably, the next time you happen to be in a similar situation to that described above you'll pay a lot of attention, maybe by considerably slowing down your movements, to be able to verify that what I've told you actually happens—and it's not clear that you're going to succeed.

The comparison between the two times in which you grasp a piece of fruit from a basket, before and after learning about the so-called *maximum grip aperture*, yields an example of the fact that "an action can be performed in different ways, with different degrees of skill, control, effort, and attention" (Bach, 1978, p. 364). In the first case, you're hardly paying attention and your movements are very fast; in the second case, you're closely following every step and, conceivably, moving a lot more slowly. Given the way in which an action is performed, which includes a complex of various degrees of skill, control, effort and attention, how much of it, if any, is part of the intention to perform the relevant action?

In the case mentioned beforehand, it should sound very reasonable that some sort of proportion relation between grip aperture and the size of the object holds. It might, for instance, strike us as natural to suppose that this proportion relation is *determined* by our intention to grasp, say, a small plum as opposed to a big orange, which we can easily visually recognize as relevantly different in size. But what exactly is the relation between the intention to grasp a small plum and the corresponding maximum grip aperture? In particular, it has to be clarified

whether the latter is (a) not part of the content of the intention to grasp the plum, and not part of the content of any intention; (b) not part of the content of the intention to grasp the plum, but part of the content of a different kind of intention (in a sense to be made more precise shortly); (c) if not part of the content of the intention to grasp the plum, part of the content of another intention (although not an intention of a different kind).

Bach (1978) seems to hold (a), as he states that

> an agent could intend to do something in a certain specified way, with a certain degree of skill, control, effort, and attention, but in general these are not matters of intention. And yet they do seem to have something to do with the agent's awareness of what he is doing and of what he is to do. However, this awareness seems below the level of intentions [...].
>
> (Bach, 1978, p. 364)

Pacherie (2003, 2006, 2008) instead supports (b), as she argues for some sort of division of labour between two kinds of intention, which she terms *present-directed intentions* (or P-intentions) and *motor intentions* (or M-intentions). Suppose that I am not entirely distracted by the conversation with another person at my table and I proceed to pick up a plum from the fruit basket as a result of a corresponding consciously formed intention. This intention is a P-intention. Grip aperture, on the other hand, is part of the content of a M-intention that specifies all the motor details of the corresponding P-intention.

I am going to develop and argue for a version of (c), by showing that where Pacherie distinguishes P-intentions from M-intentions (section 3.1 of the present chapter) there is just one kind of intention (sections 3.2-3.3 of the present

chapter). Furthermore, among the contents that Pacherie allows M-intentions to take, I'm going to exclude those contents concerning fine adjustments of actions (section 3.4 of the present chapter).

3.1 How P-intentions and M-intentions are supposed to come apart

Let us dwell for a while on the details of Pacherie's distinction. She maintains that both P-intentions and M-intentions play a role during action execution, in contrast with future-directed intentions, which are in some cases present before action onset. Both P-intentions and M-intentions, in her framework, have initiating, sustaining, guiding and monitoring functions (Pacherie, 2006, p. 149): they trigger the corresponding action (*initiating function*), sustain it until its completion (*sustaining function*, which is going to be discussed in chapter 5, section 2.1), guide the execution of the action (*guiding function*) and monitor its effects (*monitoring functions*). Pacherie's contention is that the latter two functions can be executed at two different levels. The former is consciously accessible and subject to rational constraints, whereas the latter has limited conscious access and occasionally escapes rationality constraints. P-intentions are responsible for the former kind of control, which, following Buekens, Maesen, & Vanmechelen (2001), Pacherie characterizes in terms of *rational control*, which takes two forms: *tracking control* and *collateral control*. Tracking control consists of keeping track of the progress that one is making towards her goal, trying to maximize it, while collateral control is about trying to minimize the side effects that might ensue from the action as it unfolds, to the point of abandoning the action if the side effects become too big. Both forms of rational control are exercised consciously, and are contrasted with the sort of control that is exercised at the level of M-intentions.

The notion of M-intentions stems from the idea (which has been strongly challenged—see Rossetti et al., 2003—but will be assumed only for the purposes of fleshing out Pacherie's distinction) that there exist two visual streams of information: vision for action and vision for perception (Milner & Goodale, 1995), where the latter consists in the recognition and categorization of objects. M-intentions are roughly to be identified with the motor representations built up and employed in vision for action. Three characteristics of these motor representations are highlighted by Pacherie (2006, p. 151). First, they represent object features in such a way as to enable the selection of appropriate motor patterns. For instance, if one intends to grasp an object, the characteristics of that object that are relevant for grasping will be represented—e.g., its size and shape will be represented in term of the sort of grip that they enable. Secondly, motor representations respect the biomechanical constraints governing the motor system, so that uncomfortable or awkward positions will be avoided by the motor pattern selection. Thirdly, motor representations take into account the ultimate goal of the sequence of movements in which they play a role, so that, for instance, if I intend to grasp a mug I will do so in a specific way that depends on whether I want to examine it as opposed to drink from it. What is relevant for the purposes of Pacherie's distinction is that the content of a motor representation, i.e. of an M-intention, can significantly come apart with respect to that of the corresponding P-intention. We'll see examples of this coming apart in the following.

Pacherie (2006, p. 156) reports a series of experiments (Fourneret & Jeannerod, 1998; Slachewsky et al., 2001) in which subjects' awareness of their movements was investigated. Subjects were instructed to move a stylus to a visual target. Only the stylus trajectory was visible to them in the form of a line on a

computer screen. A bias was then introduced on the line appearing on the computer screen such that there was a discrepancy between the visible line and that actually traced by the subject. The results showed that two very different strategies were employed by the subjects, depending on the amount of discrepancy between the line actually traced and that visually perceived by the subject. In particular, whenever the bias was small, subjects automatically adjusted the trajectory of their hand to it. When the bias exceeded a mean value of about 14°, subjects changed strategy and began to use conscious monitoring of their hand movements to correct for the bias and to reach the target.

The case in which subjects are unaware of the adjustments they make to the trajectory of the stylus is taken to exemplify a situation in which P-intention and M-intention come apart. When asked, subjects declared that they thought their hand had moved towards the target, but seemed to be unaware of the actual movements they had performed (Jeannerod, 2006, p. 51). Thus, their P-intention to reach the target was dissociated from the M-intention specifying the details of the hand adjustments. The way in which P-intentions and M-intentions are applied to the interpretation of these experiments is in line with the characterization of M-intentions as having a content that is not always accessible to introspection (Fourneret & Jeannerod, 1998; Slachewsky et al., 2001). A further point made by Pacherie, which is also linked to the possibility of conscious control, has to do with the timing of the action: for an action to take place quickly and smoothly, corrections need to be made at too fast a pace to be consciously accessible (Pacherie 2008, p. 188). Hence the need for M-intentions that take care of these fast and automatic corrections, in contrast with those typical of rational control that are a function of P-intentions.

So far, the seeming plausibility of Pacherie's picture, featuring P-intentions as well as M-intentions, should have been made clear. I am now going to put pressure on this picture, and in particular I'm going to challenge the idea that we need two kinds of intentions, along the lines of Pacherie's P-intentions and M-intentions, to interpret the experimental results she reports. While I agree with Pacherie that there is room for intentions even after during action performance, I disagree that more than one kind of intention is required.

3.2 How far can awareness of action performance go?

Jeannerod (2006, pp. 45-47) draws a useful distinction between being aware of having a goal, e.g., the goal of reaching a certain place, and being aware of how the goal is being reached, e.g. of the movements that my legs should perform in order to reach a certain place. One may well be aware of the former, but not of the latter—as a matter of fact, that's precisely what happens when one intends to walk to a certain place. Think again of the experimental results reported previously, and specifically of those involving moving a stylus towards a target. The target was always reached, but subjects weren't always aware of the movements they had performed in order to achieve that goal. Jeannerod (2006, p. 51) describes these situations as ones in which there is perceptual awareness without motor awareness, i.e., subjects were aware of the target of their action and were able to reach for it, but didn't know, or had partial and/or incorrect information as to how this had been possible. Question: are the awareness of having a goal and the awareness of how the goal is being reached both relevant to pick out the intention corresponding to the relevant action? If so, to what extent?

First of all, Jeannerod (2006, p. 46) points out that the extent to which one is aware of the way in which a goal is being achieved may be a function of whether the corresponding action is being learned, or whether it has been learned and is being fluently executed. When one is learning to perform a certain action, one cannot help paying more attention to the details of its execution, to the detriment of the performance. Conversely, there are occasions (e.g., piano playing) in which the only way to achieve a smooth performance is to execute one's movements with a certain degree of *lack* of awareness. It is an open question how the intention underlying these different kinds of performance, e.g. that of a pianist exercising vs. playing at a concert, should be characterized in terms of intentions, and it is unclear that Pacherie's framework provides the best way to account for these changes in the degree of conscious control. In particular, she describes M-intentions as having a content that in some cases *cannot be* accessed by introspection. But the pianist example introduces the doubt that most of the times we actually have the choice of introspecting the content of the so-called M-intentions, but avoid it insofar as doing so might hinder our performance.

We could posit just one kind of intention, which I'm going to label *motor intention*, which at once replaces P-intentions and M-intentions. The content of a motor intention can be more or less specific, and one may be more or less aware of its details, depending on the requirements of the situation. In particular, supposing that the content of a given motor intention specifies the way a goal is being reached along with the goal itself, one may only be aware of the goal, but not of how it is being reached. For instance, while on many everyday occasions it is quite possible that we don't need to pay a close attention to the fine details of our movements, there are situations—from playing an especially difficult passage on

the piano to verifying the existence of the maximum grip aperture phenomenon—in which we may be aware of very fine details of action execution, which pertain to how the goal is being reached.

3.3 A proposal: one kind of intention, many possible content/consciousness specifications

As we've seen, accuracy of performance often requires that we dramatically limit our conscious control of an action. Pacherie expresses this idea by resorting to the notion of the *tempo* of an action (e.g., 2006, pp. 150-151), which is to be understood as the pace at which an action unfolds. A slow tempo is one which offers more possibilities for conscious control, insofar as one has more time to decide on the details of action execution, and on the modifications to make on the course of action. If one is practising a piano piece, without meaning to attain the pace at which she'll actually perform it during a concert, she may for instance take the time to calibrate the movements of each finger in those passages that are especially tricky. A fast tempo is one that considerably limits the possibilities of conscious control, insofar as little time is conceded to make the necessary adaptations to accomplish a given action. Pacherie (2006, p. 150) gives the example of a game of tennis, in which one is allowed little time to decide how she's going to return a serve. In the pianist example, a concert performance may be characterized by a fast tempo: the rhythm at which the piece is to be played constrains the amount of conscious control that can be devoted to the execution to a point at which it's very limited. Pacherie's contention is that two different kinds of intentions are required to deal with these differences in tempo: specifically, while an M-intention always exercises control over an action,

> for a P-intention to play its role of guidance and control, it must be the case that the tempo of the action is not faster then the tempo of conscious rational thought; or, more accurately, it is only on those aspects of an action the tempo of which does not exceed the tempo of conscious rational thought that P-intentions can have rational control.
>
> (Pacherie, 2006, pp. 150-151)

According to Pacherie, then, the difference between the pianist at the practising stage and at the performance stage should lie in something like the following: while in the former situation she may have a series of P-intentions such as "alternate index, ring, middle and ring finger" (while playing a trill), in the latter case all these P-intentions will be no longer present and her performance will solely be guided by the relevant M-intentions. Is this an effective description of the events in question? After all, the boundaries between the two kinds of situation may not be so rigid. Even during a concert performance, a pianist might be aware that there is a certain point in the piece that she has to pay special attention to, and thus exercise as much rational control on that point as the rhythm of the execution allows. In Pacherie's framework, this situation would be described as one in which every now and then the relevant P-intentions flank the ongoing M-intentions. Is it necessary to posit two kinds of intentions to characterize this situation?

Velleman (2007, pp. 213-214), for one, would probably judge the piano concert situation as one in which intentions are most likely to get in the way, and are best left aside. He resorts to the work of the psychologist Mihaly Csikszentmihaly (1990, 1997), who identified some activities as able to provide a sort of experience which he terms of *flow*. These are activities demanding the

exercise of appropriate skills, and in which people become so involved that the activities become spontaneous, almost automatic (Csikszentmihaly, 1990, p. 53; quoted in Velleman, 2007, p. 214). Supporting Csikszentmihaly, Velleman insists that if we are to find flow we have to suspend deliberation and planning, and thereby to suspend intentions. As Velleman doesn't embrace any equivalent of Pacherie's notion of M-intention, he altogether rules out the involvement of intentions in those activities that generate the experience of flow. He and Pacherie clearly agree that *something* is lacking in this kind of activities, but how should this something be characterized?

First, a clarification remark is in order. Both in the concert and in the practice situation, the pianist *intends to do something*—playing, say, Beethoven's Moonlight sonata. One may object at this point that even this overarching intention differs in the two cases, for in the former one intends to *execute* Beethoven's Moonlight sonata, and to *practise* it in the latter, but, assuming that this subtlety has any substance, let us ignore it for the sake of argument. Given the similarity in terms of overarching intention, how is the lack pointed out by Pacherie and Velleman alike to be characterized? One possibility is that the content of the intention is much richer in the practice case with respect to the concert situation, for in the practice case the intention to play Beethoven's Moonlight sonata can be re-described as the intention to lift alternatively one finger and then another finger and so forth, whereas in the concert case this re-description is not available. This solution assumes that what's consciously attended to has a bearing on the specification of the content of the relevant intention. In the case in which the pianist devotes more or less conscious introspection to his performance during a

concert, this solution has it that the content of his intention expands or contracts so as to include more or less details.

Another interpretation is possible, and it is independent of the content specification. According to this interpretation, the two situations differ in terms of the amount of conscious introspection devoted to the activity by the subject, but this difference is orthogonal to the question of what kinds of intentions, if any, are involved. It could be the case that the two situations do not differ at all in terms of intention and its content: in both, it might be right to say that the pianist intends to lift first one finger, then another, and so on—that is, the richest possible characterization of the content of her intention may hold of both cases, while what's making a difference is the degree of conscious introspection. I do not see a straightforward way to adjudicate between the two interpretations[43], but the point is that neither requires the positing of two kinds of intention. The second is silent on how many kinds of intentions are involved, and the former simply requires a greater or lesser specification of the content of an intention, where no more than a kind of intention is needed.

The example of the pianist which occupies this section focuses on the presence or absence of introspection, without positing any contrast between the contents accessed via introspection and those belonging Pacherie's M-intentions. This may be thought of as an easy case to handle without a distinction between P-intentions and M-intentions. What about those cases in which conscious control

[43] In chapter 6 (section 5) I'm going to pronounce myself as to how many intentions and what kinds of intentions are involved in a number of examples. Piano playing, however, is of a much higher level of complexity with respect to the actions that I'm going to consider in chapter 6, so I'm not ready to favour any of the two interpretations just put forward in the present case.

and motor control seem to conflict to some extent? Do those require a distinction between two kinds of intentions? I am going to argue that this is not the case.

3.4 Intention and fine-tuning of action: two of a kind?

According to Pacherie (2006, p. 154), M-intentions are also involved in the adjustments to our motor commands that are employed in the generation of our actions. She appeals to a series of experiments to show that we are often unaware of those adjustments, even though they may end up in contrast with our own actions. Some pointing experiments (Goodale et al., 1986; Castiello et al., 1991) are based on a task in which subjects are required to point to a target with their finger. The results show that subjects can perform the task accurately on all trials, including those in which the target is suddenly displaced by a few degrees and the pointing trajectories are adjusted in accordance with this displacement. More to the point, subjects are unaware of this displacement. Pacherie interprets these and other results in the same vein in terms of the fact that M-intentions, which are supposed to underlie the adjustments of which subjects remain unaware, have a dynamics of their own, which is not entirely under the control of P-intentions, such as the P-intention to point at a target.

Now, why think that it is the function of intentions (though of a specific kind) to provide these adjustments? Pacherie's answer to this question should be something along the following lines: since guidance and control are among the functions of intentions at the time of action execution, and fast movement corrections serve the purposes of guidance and control of action, then they should be described in terms of intention. But here is a problem, which should lead us to reconsider, or at least rephrase, the second premise.

Movement corrections such as those featuring in the experiments reported are pervasive in the course of our voluntary movements, and do not necessarily play a role in action. Consider a case in which we're idly listening to a speech while standing up still in the middle of a crowd. If we have any intentions at that point, they'll have to do with listening to the speech, and will be quite independent of our bodily position. The intention to listen to the speech might prompt the intention to get closer to the speaker, or some such, but the fact that we're in a specific bodily position, i.e. standing up, is completely irrelevant to the fulfilment of our intention. In particular, therefore, at that point we have no intention to stand up, nor can we say that the action of standing up is unfolding. Yet there is evidence (see, e.g., Ramenzoni et al., 2011) to the effect that even standing up requires a series of continuous small and imperceptible postural adjustments. The above example is meant to show that such fast corrections don't necessarily serve the purpose of guiding and control of an action, since there is no action involved, at least not to do with the person's standing up.

The upshot is that, on pains of including too much within the content of an intention, a promising interpretation of the pointing experiments might be in terms of just one kind of intention plus a series of postural adjustments. It is plausible that these adjustments do not have anything to do with intentions, at least below a certain threshold. Think of the experiments (Fourneret & Jeannerod, 1998; Slachewsky et al., 2001; reported in Pacherie, 2006, p. 156) in which subjects were instructed to move a stylus to a visual target. Only the stylus trajectory was visible to them in the form of a line on a computer screen. A bias was then introduced on the line appearing on the computer screen such that there was a discrepancy between the visible line and that actually traced by the subject.

The results showed that, whenever the bias was small, subjects automatically adjusted the trajectory of their hand to it. When the bias exceeded a mean value of about 14°, subjects began to consciously monitor their hand movements to correct for the bias and to reach the target.

Pacherie would say that a P-intention to move the stylus towards the target is flanked by a M-intention which takes care of the postural adjustments that, below the threshold of 14°, are not consciously accessed. But, in the light of the considerations based on the work of Ramenzoni et al. (2001), an alternative interpretation would have it that, when the bias is below 14°, only one intention, i.e. the intention to move the stylus towards the target, is in play, while a series of postural adjustments are performed in conjunction with it to help fulfilling that intention. In the case in which the bias exceeds 14°, on the basis of the previous reflections we could either say that the content of that intention gets supplemented by the further intentions to adjust one's movements, or that the content of the intention remains the same, but more introspection is employed during action execution. None of these cases requires the positing of a further kind of intention.

In support of the proposed interpretation according to which intentions and fine-tuning of action may come apart, consider the following experimental result. Graziano and Aflalo (2007, reported in Uithol et al., 2012, p. 1082) carried out a study in which the premotor areas of a macaque monkey brain was stimulated for a relatively long time (500-1000ms). This stimulation managed to evoke complex sequences of movements directed to a given location. An example of this sequence would be given by grasping, bringing to the mouth, orienting the head towards the hand and opening the mouth. The striking characteristic of these movements is

that they didin't display any flexibility: in particular, when an arm encountered an obstacle on its way to bringing something to the mouth, it didn't correct its trajectory and thus ended up blocked against the obstacle (Graziano, 2010, p. 461, reported in Uithol et al., 2012, p. 1082).

Conclusion

In this chapter I've introduced the notion of motor representation, and shown to what extent it bears a significant similarity to that of intention. This has given me the chance to introduce also the notion of motor intention. Motor intentions are not a separate kind of intentions with respect to the ones we know and love (examined in the course of chapter 2), but are characterized by a specific kind of content, which is relevantly similar to that of corresponding motor representations. Recalling a notion that was introduced towards the end of the first chapter (section 5), motor intentions represent motor outcomes. The relation holding between motor intentions and motor representations, which is going to be made more precise in the course of the following chapter, paves the way for an answer to the question as to how intentions connect with actions.

I've suggested that motor intentions can do the work of both Pacherie's P-intentions and M-intentions, by insisting on the idea that any kind of content that may be even temporarily introspected—including contents that were classified as belonging to M-intentions—can be the content of an intention standardly conceived. Now, where does this leave us with respect to the intention to grasp a plum by stretching one's fingers in accordance with the maximum grip aperture phenomenon? I claim that this is an empirical question. The present chapter has established that the intention to grasp a plum is a motor intention, which is a

standard kind of intention characterized by a commonality in content with respect to the relevant kind of motor representation, or motor outcome. Since, as mentioned in the first chapter (section 3.1), it is likely that some motor representations involved in the act of grasping are going to specify in their content the extent to which fingers open. While in principle the content of these representations could be relevantly similar to that of a corresponding motor intention, it is an empirical question whether any motor intention specifies that kind of content, and an answer to this question partially hinges on the question as to whether the maximum grip aperture phenomenon is one that we're capable of introspecting, thus rendering it apt to be the content of an intention. This issue is going to be tackled in the next chapter.

§4. What is the content of a motor intention?

Introduction

In the previous chapter, the notion of *motor intention* was introduced, namely that of an intention whose content is relevantly similar to that of an motor representation. It was also explained that it is possible for us to gather insight into the content of a motor representation by means of *motor images*, where a motor image is a mental image of the self performing an action. These images preserve many features of the imagined actions—for instance, a motor image takes roughly the same time to unfold as the imagined action. It would then seem that motor images represent imagined actions in great detail. If we rely on motor images in order to draw conclusions about the content of motor representations, we may think that all motor representations represent actions in as much detail. If the content of motor intentions is thought to be relevantly similar to that of a corresponding motor representation, we may wonder what consequences this has for how the content of motor intentions should be characterized. In particular, is this a kind of pictorial content or is it propositional? Is it conceptual? If we thought that motor intentions had a pictorial kind of content, impossible to render in propositional terms or to describe in terms of concepts (or in concepts available to the average subject), then it would become hard to justify calling them intentions.

The present chapter is going to answer the aforementioned questions by shedding light on the variety of motor representations, and on their different levels

of specificity. In particular, I'm going to resort to the notion of an action hierarchy (Shallice, 1988; Jeannerod, 1994; Hamilton & Grafton, 2007; Csibra, 2007), in order to show that an action can be represented in greater or lesser detail depending on the hierarchical level under consideration. On the basis of this variety, it'll be pointed out that, while all motor intentions have something in common insofar as their content involves sequences of bodily configurations, only some motor intentions may be safely held to have a propositional content that is conceptual under the most demanding construal of the notion.

1. Are motor intentions really intentions?

In the previous chapter, along the lines of Jeannerod (1994, 2006), I introduced the idea that the content of a motor image may be a good guide to the content of a corresponding motor representation, and I described motor intentions as having a content that is relevantly similar to that of the corresponding motor representations. In this section I'm going to point out potential issues related to the similarity between motor intentions and motor images.

Some difficulties arise from the fact that motor imagery is part of the broader phenomenon of mental imagery. Within cognitive science, a long-standing debate (the imagery debate) has been concerned with how the format of mental images should be characterized. One side of the debate has it that mental imagery essentially differs in format with respect to linguistic propositions (Paivio 1986, referenced in Jeannerod, 1994, p. 188). In particular, according to this side of the debate, mental images have a *pictorial* or *analog* format (Kosslyn et al., 1979, referenced in Jeannerod, 1994, p. 188). Mental images, that is, exhibit spatial representational properties of the same sort as those possessed by pictures—for

instance, any part of a picture represents a part of the depicted scene, and spatial relationship among objects in the depicted scene are rendered via spatial relationships on the picture surface (see Thomas, 2010). The other side of the imagery debate, on the contrary, has it that mental imagery does not differ in format from other forms of mental representation: they all have the same abstract kind of representational format (Pylyshyn, 1984, referenced in Jeannerod, 1994, p. 188). While this is not the place to try and adjudicate the mental imagery debate, let us suppose for the sake of argument that the pictorial side of the debate is right. What bearing does this have on the notion of motor intention?

Arguably, intentions have a propositional format, insofar as they are standardly conceived as propositional attitudes (see, e.g., Bratman, 1987) and so it's difficult to see how there could be commonality in content between any kind of intentions on the one hand and motor images on the other hand, if the latter are characterized as having a pictorial format (see Butterfill & Sinigaglia, 2012, on the difference in format between intentions and motor representations). There are further difficulties connected with the idea that motor imagery is an instance of mental imagery: consider the notion that there is a gap between the expressive power of a picture compared with that of a propositional description. Imagistic representations are often conceived to be analog insofar as the properties they represent vary continuously as opposed to propositional representations, which are digital insofar as what they represent varies in a discrete fashion (see Dretske, 1981, on the analog vs. digital distinction; Pitt, 2012). Granting all these things, how can the content of an intention ever capture that of a motor image?

As if this wasn't enough, there is an additional difficulty connected with the fact that motor imagery is special with respect to mental imagery generally

conceived. As Jeannerod (1994, p. 189) points out, motor imagery is a first-person process in which the action is represented kinaesthetically, and not so much visually (although, arguably, a visual component may well be part of a motor image—after all, even from a first-person point of view, whenever I act my own body looks a certain way as well as feeling in a certain way). Jeannerod draws a contrast between the difficulty in verbalizing motor images and the relative accessibility of visual images. Putting all these difficulties together, how can we justify the idea that motor intentions are intentions?

2. Action hierarchies and variable content specification

Consider the following example (from Csibra, 2007, pp. 439-440) of an action hierarchy: I want to eat an apple, and this result takes a few steps to be achieved, each of which may be carried out in different ways. For instance, at some point the apple needs to be brought in contact with the mouth. In order to obtain this result, I may either bring the apple to the mouth or move the mouth to the apple. Suppose I choose the former way to bring the apple in contact with the mouth. Once again, there's room for performing this action in different ways, insofar as I may grasp the apple by means of a whole-hand prehension, or pick it up by the petiole by means of a precision grip. Suppose I decide to grasp the apple with a whole-hand prehension. So far, I've highlighted three levels of an action hierarchy: eating the apple occupies the top one, the two ways of bringing the apple in contact with the mouth are at a level below, and the two ways of grasping the apple at an even lower level. It should be relatively uncontroversial that any token action may be described at different levels of a hierarchy along the above lines (as a matter of fact, this idea has been extensively examined in chapter 1).

A fundamental characteristic of action hierarchies as described above is that, as we climb down the hierarchy, we find that the action is specified in increasingly great detail (Jeannerod, 1994; Hamilton & Grafton, 2007; Csibra, 2007; this notion is going to be made more precise in chapter 6). If comparisons between the motor image and the corresponding action in terms of timing, biomechanical constraints and so on (see previous chapter and Jeannerod, 1994; 2006, pp. 23-ff.) are to be possible at all, the content of a motor image has to be as detailed as to encompass even the finest level of the hierarchy described in the above example. In particular, it's likely that the motor image of my grasping an apple will take into account, if not the kind of grip employed, at least the fact that I grasp the apple with one of my hands. Therefore, motor representations whose content is relevantly similar to that of a motor image are going to encompass roughly as much detail as the corresponding motor image. But this kind of representations does not exhaust the variety of motor representations that are seemingly at work in action production (and observation, although the latter won't be tackled in the present work).

Jeannerod (1994, p. 200-201), who has extensively described the relation between motor representations and motor images, acknowledges that detailed motor representations along the above lines do not exhaust the representation mechanisms that operate in action production, insofar as motor representations at the lower levels need to be embedded in broader encompassing motor representations which represent long-term goals. This in principle need for motor representations higher up in the hierarchy is complemented by experimental evidence concerning the representation of actions at a higher degree of generality with respect to the detailed representations linked to motor images. To go back to

the apple grasping example, note that, among mirror neurons (see chapter 1 for a general introduction), some have been recorded (Gallese et al., 1996) that fire only in the case of *grasping with a precision grip* or of *grasping with a whole-hand prehension*. These neurons, that is, respond selectively for one of the two grasping options highlighted in the apple grasping example, and are called *strictly congruent mirror neurons*. In addition to these, neurons have been recorded that fire in correlation with grasping with the hand, independently of the kind of grip. These have been termed *broadly congruent mirror neurons*, and their level of specification of the action corresponds to the middle level in the apple grasping example. Broadly congruent mirror neurons can be considered evidence for the existence of motor representations at a higher level of generality.

Experimental results concerning mirror neurons provide evidence for motor representations representing actions at even higher levels of generality, insofar as (see chapter 1, section 2) there exist neurons firing in correlation with grasping not only regardless of the kind of grip, but also regardless of the effector employed (so, irrespectively of whether something is being grasped with the hand or with the mouth), and even in correlation with grasping carried out by means of tools such as pliers. What is crucial for our purposes is that these neurons are evidence for the representation of the action of grasping at a very high level of generality. Now, given that this level of generality is such as to leave open whether the subject is moving her mouth or hand, and she's moving her hand in what way, how does this sit with the similarity in content between a motor image and an motor representation? The issue is that if I have to form an image of myself grasping, this image can't leave it unspecified whether I'm grasping with my hand or with my mouth. Even if the image in question is kinaesthetic in character and

not visual, it'll still have to specify my action in one of these two ways. In the light of this, are there any *specific* motor images whose content may be relevant for the specification of the content of a grasping motor representation? And, if not, what does this tell us about the content of motor representations at a higher degree of generality, and, more to the point, about the content of the corresponding motor intentions?

3. Is the content of motor intentions conceptual?

The piece of reasoning carried out in the earlier section should lead us to doubt that, when it comes to motor representations of higher level outcomes such as grasping (as opposed to, e.g., grasping with a precision grip), these representations are still going to be relevantly similar to motor images. The higher degree of generality that they enjoy may lead us to think that a relevant change in format occurs such as to make them amenable to be rendered in propositional terms (*contra* Butterfill & Sinigaglia, 2012). Not enough is known in order to be able to conclude that the motor representation of grasping has the right kind of format to be integrated in a piece of practical reasoning (it is still possible that its content consists in a motor image corresponding to what the subject considers a prototypical act of grasping), but that is not crucial for my purposes. What matters and should be relatively uncontroversial is that intentions involving contents of such a high level of generality as 'grasping' exist. And it seems that there may indeed be pieces of practical reasoning that incorporate intentions with that kind of content. Consider something along the lines of:

(1) I intend to get hold of the umbrella.

(2) In order to get hold of an umbrella I have to grasp it, so

(3) I intend to grasp the umbrella.

To sum up, the higher degree of generality of certain motor representations, regardless of their actual format, warrants conceiving the corresponding motor intentions as propositional attitudes. Furthermore, it shouldn't be difficult to see that they have a conceptual content, as the characterization of an action in terms of GRASPING is available to the average thinker[44]. Can the same be said about the motor intentions whose content is relevantly similar to that of the corresponding motor images? Or should they be classified as having a nonconceptual content?

The notion of nonconceptual content (Evans, 1982) has a bearing on the present discussion in at least two ways. One has to do with the idea that, as human beings have limited conceptual capacities, we may simply not have enough concepts to capture the content of our motor intentions. Secondly, even when there are concepts for describing the content of at least some motor intentions, they might not be the kind of concepts that the average thinker possesses. Since we would expect the content of intentions, *qua* propositional attitudes, to be conceptual, let us see whether and when we need to conceive the content of motor intentions as nonconceptual.

As for the first issue, it has been argued by Evans that the content of perceptual experience outstrips our capacities for conceptualization:

[44] Interestingly, it has been proposed that the activity of mirror neurons that fire in correlation with the various instances of grasping may underlie the concept GRASPING (see Gallese, 2003; Sperber, 2004; Jacob, 2009).

> Do we really understand the proposal that we have as many concepts as there are shades of color that we can sensibly discriminate?
> (Evans, 1982, p. 289)

An analogous problem may be extended from perceptual experience to motor intentions (something that Pacherie, 2011, has considered), insofar as the content of motor intentions may be thought to outstrip our conceptual capacities. Now, notice that I previously highlighted the existence of different kinds of motor intentions, some of which have a kind of content that is relevantly similar to that of corresponding motor images, and some that don't, being much more abstract. The latter kind of motor intentions have such a high level of generality that there doesn't seem to be a problem with characterizing them in conceptual terms: indeed, concepts such as GRASPING characterize this kind of intentions well enough.

Things are not so simple when it comes to motor intentions at the lower levels of an action hierarchy. In section 2 I mentioned the digital vs. analog distinction in order to point out that there is a potential difficulty residing in the content of a motor intention being significantly like the content of a motor image. The difficulty consists in the fact that the content of a motor image may outstrip our capacities for conceptualization, so that, at least in the case of those motor intentions whose content is relevantly similar to that of a motor image, it seems that those intentions must have a nonconceptual content. There are two possible answers to this point. One has been given by Pacherie (2011): she holds that the motor representations involved in action initiation and control are nonconceptual (on the basis of a version of the priority argument), while holding that these nonconceptual representations ground what she refers to as *executable concepts* of

actions. As soon as one masters a given action, and is capable of initiating it at will (by contrast with apraxic patients of individuals affected by utilization behaviour), one can thereby be said to possess the related executable concept. Another answer can be given along the lines of Butterfill and Sinigaglia (2012), on the basis of McDowell's (1994) response to the earlier mentioned piece of reasoning by Evans. McDowell holds that we may well capture the content of our perceptual experience with whatever fineness of grain by means of what he calls *demonstrative concepts*, such as 'that shade' (to keep to the colour example). In particular, if we express the content of a motor intention by means of a demonstrative which makes reference to the relevant detailed motor representation, the content of the motor intention in question can be characterized as conceptual.

Both strategies rely on assumptions that one may or may not be ready to accept—on the legitimacy of talk of executable concepts and of demonstrative concepts, respectively. If neither is accepted, this is going to mean that the variety of motor intentions includes some intentions whose content is nonconceptual, for which some may therefore wish to withdraw the term *intention*. What should however be clear from the earlier section is that the variety of motor intentions still includes intentions, which I characterized as occupying higher levels of an action hierarchy, that uncontroversially have a conceptual content.

Now let's consider a second respect in which the notion of nonconceptual content becomes relevant for motor intentions. Recall that in the previous section I mentioned the existence of neurons selectively firing in response to precision grip vs. whole-hand prehension. While the precision grip and whole-hand prehension movements are captured by the concepts bearing the same name, it is also true that these concepts are not possessed by the average thinker. This may be

considered enough to say that whenever someone (who is not a neurophysiologist or an experimental psychologist) intends to carefully grasp an apple by the petiole, by explicitly intending to perform the movements that are technically characterized as forming a precision grip, she thereby possesses a motor intention with nonconceptual content. Answers to this kind of worry may be along the same lines as those provided in response to the earlier worry. Once more, depending on the assumptions that we are ready to accept, the content of the intention to grasp using a precision grip may be thought to have a conceptual or nonconceptual content. Again, even if we had to conclude that intentions such as that of grasping something with a precision grip are better described as having a nonconceptual kind of content, thus leading some to withhold the term 'intention' for them, this still leaves room for motor intentions, such as the intention to grasp, whose content is uncontroversially conceptual.

In the course of the preceding sections, the possibility to subdivide motor representations into two categories has emerged. On the one hand, there are motor representations lower down the hierarchy, which have a very detailed kind of content, likely to be constituted by a motor image and to be nonconceptual. On the other hand, there are motor representations higher up the hierarchy, that have a very abstract kind of content, which could in principle be rendered in propositional terms. On the basis of this subdivision, we're now in a position to specify what the similarity relation between motor intentions and motor representations consists in, thus shedding light on how intentions connect with actions. Along the lines of Butterfill & Singaglia (2012), I'm going to propose that motor intentions refer to action outcomes of the more detailed kind by *deferring* to the corresponding motor representation. When it comes to motor representations

of a more abstract kind, however, if their content can indeed be rendered in propositional term, then that content is going to be *identical* with that of the corresponding motor intention.

4. What is the content of motor intentions?

Now that I've described the variety of motor intentions and their potential differences in terms of content, I'll focus on what brings them together. I'm going to take the intention to grasp as a paradigmatic example of a motor intention. The intention to grasp has as its content a state of affairs that crucially involves a specific *sequence of bodily configurations.* To fulfil my intention to grasp, say, a peanut, I must end up – in the most ordinary of cases – with the peanut between my fingers. I mention fingers since they are the bodily parts most likely to be employed to this end, but it is possible to imagine a situation in which my hands are occupied, or are somehow blocked, or do not exist at all (as may be the case with aplasics, subjects who congenitally lack one or several limbs), and I have to grasp the peanut with my mouth or with my feet, where this is possible. Not *any* bodily part will do (I can't grasp anything with my nose, due to anatomical constraints), but *some* bodily part has to be employed. The important feature, then, is that the content of my intention crucially involves one of a *limited set of bodily parts*, which has to undergo a specific change, one, for instance, that brings it in a specific relation to a certain object.

At this point, one may be tempted to define particular intentions of this kind (e.g., the intention to grasp something, or the intention to tear something) exclusively in terms of the possible final states that count as the accomplishment of that particular intention. For instance, in the case of catching, one may believe it

sufficient to characterize the content of the intention to catch a ball in terms of the ball being in my hands. But that's not so, for that state of affairs won't count as the satisfaction of my intention to catch a ball if it is the result of someone's carefully placing the ball between my hands (Pacherie, ms.). That's simply not catching, for the notion of catching implies, for instance, that the object to be caught is not under my (or anyone else's) control before I catch it. The final state by itself doesn't suffice to characterize what counts as the realization of this kind of intention: hence the need of the aforementioned notion of a sequence of bodily configurations. That sequence is apt to characterize the relevant kind of outcome, which is not simply a state of affairs, but rather an action. Intentions whose content crucially involves a sequence of bodily configurations will be what we term *motor intentions*. Examples are the intention to grasp a peanut and the intention to catch a ball.

To clarify, compare these intentions with, e.g., the intention to be good, or the intention to improve my grades. In both these cases, no specific bodily part is involved in the accomplishment of my intention, nor any sequence of bodily movements. Clearly, neither of them is a motor intention. Not even the intention to go to the cinema counts as a motor intention. Although the accomplishment of this intention does involve a displacement of some sort (if I am already in the cinema, I can at best intend to remain there, not to go to the cinema, unless it's about a different one), it doesn't imply any specific sequence of bodily movements. I may get to the cinema in any of the following ways: on foot, by car, by bus, or even, in a sufficiently fanciful scenario, by teletransportation, a case that allows one not perform any bodily movement at all. So motor intentions are intentions of a specific kind, such that their content essentially refers to bodily configurations. In

this sense, what they represent is sufficiently close to motor instructions. But, given the level of abstraction that at least some motor intentions enjoy with respect to motor instructions, the former are also significantly different from the latter, insofar as the former, but not the latter, represent action outcomes (those which, in chapter 1, section 5, I've termed *motor outcomes*) with a certain degree of abstraction.

At this point, one might raise the doubt that motor intentions are still too unlike standardly conceived intentions in that motor intentions are more apt to account for spontaneous, unreflected actions, as opposed to actions planned in advance. In short, one might raise that doubt that motor intentions provide the best interpretation for Searle's intentions in action, while retaining the idea that intentions standardly conceived are much closer to what Searle calls prior intentions, and that the two are significantly different. The following chapter is going to show that Searle's prior intentions and intentions lie on a continuum, so that they're best accounted for in terms of just one kind of intention. I'm going to use the notion of motor intention as the best and most plausible way in which Searle's intentions in action be interpreted (see Pacherie, 2000, for a similar strategy).

Conclusion

In the present chapter, I've highlighted the existence of a variety of motor intentions, corresponding to different levels of an action hierarchy. While the motor intentions corresponding to the lower levels of the hierarchy bear a significant resemblance to motor images in terms of content, the content of motor intentions higher up in the hierarchy is not so straightforwardly identifiable with

that of a motor image. This implies that motor intentions higher up in the hierarchy are less controversially classified as intentions, insofar as they have a propositional format and a conceptual content. As far as motor intentions lower in the hierarchy, the acknowledgement that they are propositional attitudes, and of their having a conceptual content relies on the acceptance of the demonstrative strategy and/or of the notion of executable concepts. What brings together all these different kinds of motor representation is the fact that their content involves reference to sequences of bodily configurations. Now I'm in the position to specify the similarity in terms of content between motor intentions and motor representations, introduced in chapter 3, as follows. Motor intentions at the lower levels of the hierarchy refer to action outcomes by *deferring* to the relevant motor representations (see Butterfill & Sinigaglia, 2012) whose content is likely to be a motor image; motor intentions at the higher levels of the hierarchy are likely to have the *same content* as that of corresponding motor representations, the latter being apt to be rendered in propositional terms.

§5. Prior intentions vs. intentions in action: exploding the myth

Introduction

"Intention is Janus-faced, tied both to intentional action and coordinating plans", writes Michael Bratman (1984, p. 404). It is part of our ordinary conception that intentions and intentional actions are not unrelated, and yet it is hard to specify what their relation is. Searle (1983) has put forward an account, featuring prior intentions and intentions in action, by means of which he proposes to reconcile the two faces of intention. This and analogous attempts have been criticized by Bratman (1984, 1987) as expressions of what he terms *a Simple View* on the relation between intentions and intentional actions.

I am going to propose a unitary view of intention that evens out the differences between intentions in action and prior intentions. My view will incorporate Searle's suggestion that all intentional actions should be explained in terms of intentions. This claim is by no means obvious, and has been variously opposed (e.g., by Bach, 1978, as well as by Bratman, 1987). To reach this unitary standpoint, I shall first of all expound Searle's distinction between prior intentions and intentions in action, and I will subsequently suggest that we should see no distinction at all between the two.

1. The alleged differences between prior intentions and intentions in action: temporal characteristics and content

The assumption that Searle works on is that intentions, just like beliefs, are Intentional states, namely mental states characterized by a directedness upon

objects and states of affairs in the world (1983, p. 1) – whenever I intend, there must always be something that I intend to do. Intentions have a representational (or Intentional) content and a world-to-mind direction of fit. For an intention to be fulfilled, the world has to conform to it. On these premises, Searle proceeds to draw the distinction between prior intentions and intentions in action.

Searle (*ibid*., p. 84) points out that, while some actions are preceded by the formation of the intention to perform them – e.g., the intention to kill my uncle, formed (say) one day before the deed – there are some rather spontaneous actions, such as hitting someone all of a sudden, that are not preceded by any intention to so act. On these assumptions, Searle introduces the distinction between *prior intentions* and *intentions in action*, the first characterization of which is in temporal terms. Prior intentions are those formed prior to the initiation of an action, as opposed to intentions in action, which are present as an intentional action unfolds. An intention in action is the component that makes an action intentional, regardless of whether the latter has been preceded by an intention to so act.

The following worry might at this point arise: by introducing intentions in action, Searle acknowledges all intentional actions, including spontaneous actions, as accountable for in terms of intentions. As we will see by looking at the content characteristic of intentions in action, Searle defines these intentions in such a way that they do indeed turn out to be present in all intentional actions. Still, it is a moot point whether all intentional actions, and specifically spontaneous actions, can be explained in terms of intentions (Bratman, 1987, p. 126, for instance, explicitly denies this). So why does Searle allow that intentional actions should be explained in terms of intentions? On the basis of his views about Intentionality, he notes that the conditions of satisfaction of intentions are intentional actions. Then

he suggests as a provisional (but, according to him, on the right track) view that "an intentional action is simply the conditions of satisfaction of an intention" (1983, p. 80), and thus makes room for intentions underlying all sorts of intentional actions by introducing seemingly *ad hoc* intentions.

So, one might be tempted to get off the bus already at this point and to reject his account, or any accounts based on his, on the grounds of this move. I believe that most intentional actions, including spontaneous actions, can be accounted for in terms of intentions, but the reasons why this is so will have to wait until the following sections. Until that point, in spite of the worries that one might have already at this stage, let us assume for the sake of argument that Searle's introduction of intentions in action is justified.

Let us now turn to the differences between prior intentions and intentions in action in terms of content. For the time being, following Searle, I will keep to very simple intentions such as the intention to raise my arm. As highlighted by Pacherie (2000, p. 405), there are three related aspects concerning content in which the two kinds of intention differ, and they are summarized in the following passage:

> [...] the contents of the prior intention and the intention in action look quite different, because [...] the prior intention represents the whole action as the rest of its conditions of satisfaction, but the intention in action presents, but does not *re*present, the physical movement and not the whole action as the rest of its conditions of satisfaction. [...] Another difference is that in any real-life situation the intention in action will be much more determined than the prior intention, it will include not only that my arm goes up but that it goes up in a certain way and at a certain speed, etc.

(Searle, 1983, p. 93)

A first difference is in terms of Intentional content: while prior intentions have the whole action (I perform the action of raising my arm, in the adopted example) in their conditions of satisfaction, intentions in action have the physical movement (that my arm goes up) in their conditions of satisfaction. A second difference has to do with the contrast between presentations and representations, introduced by Searle (1983, p. 46) within the context of the Intentionality of perceptual experiences. The idea he puts forward is that the perceptual experience of an object has a kind of directedness, immediacy and involuntariness that the belief about the object in its absence lacks. The perception, so to speak, imposes itself on the perceiver both vividly and unavoidably. The same difference between presentation and representation is said by Searle to apply to the contrast between the perception of an object and the memory of such perception. This analogy is exploited to characterize the contrast between the prior intention to raise my arm and my intention in action to raise my arm, the formal relations among which are the same as those holding between (respectively) the visual memory of a flower and the visual perception of a flower. The third difference concerns the fact that the content of an intention in action is much more determined, or, we might say, much more fine-grained, than that of a prior intention. In the case of the arm raising example, the prior intention might leave unspecified a series of kinematic and dynamical details that the intention in action does specify, e.g. the path followed by the arm rising, the speed at which it rises, etc.

As Searle himself notes (*ibid*., p. 94), since the contents of the two kinds of intentions have been characterized as very different, the question arises as to how

prior intentions and intentions in action are related to each other. Searle puts forward the idea that, whenever both a prior intention and the intention in action to perform a specific action are present, the prior intention causes the intention in action, which in turn causes the physical movement. The action is constituted by the intention in action together with the physical movement. By transitivity, the prior intention causes the physical movement.

At this point, the distinction between prior intentions and intentions in action has been presented both in its temporal aspects and in those concerning content. One may still wonder, though, what should lead us to accept that, for such simple intentions as that to raise one's arm, there should be two distinct kinds of intentions at work. There is at least a way to understand the distinction between prior intentions and intentions in action, which I'm now going to expound, such that all the highlighted differences fall into place in an at least prima facie natural way.

Suppose one is about to perform an intentional action, e.g., one is about to raise one's arm, and has the intention to do so prior to the performance of this action. Before action onset, one has a so-called prior intention to raise one's arm at a certain point in the future. The content of the prior intention features a deeply unspecified template of the action that one intends to perform: one may simply intend to raise one's arm, following no matter what trajectory, no matter at what speed – hence the representational (vs. presentational) character, and the lack of specification (e.g., the failure to specify kinematic and dynamical details).

By contrast, once the time for acting comes, an unspecified template of the action one wants to perform no longer suffices to fill the content of an intention in action. On the contrary, the sequence of physical movements that realize this

action has to be specified in detail. It follows that the intention in action will have a presentational character, since at the time of action one is actually executing, and no longer just planning, a series of movements, and is in touch with those movements in an immediate and vivid way that is supposedly unavailable at the time of in advance planning. Furthermore, no detail of the physical movements that realize the action, be it kinematic, dynamical or whatever, can be left unspecified. One's arm will be made to rise following a specific trajectory, at a determined speed. Now that I've drawn this seemingly plausible picture, I'll proceed to show in the following section why we shouldn't accept it.

2. Annihilating the differences, exploding the myth

2.1 Temporal differences

So far, I've described Searle's distinction between prior intentions and intentions in action as based on temporal characteristics as well as on content-related features. In what follows, I will show that such alleged differences can be annihilated, starting with the temporal characteristics. The question is: at which point in time, if any, do prior intentions expire? I.e., should the fact that a prior intention causes an intention in action, which is part of the action, be interpreted as meaning that prior intentions expire at action onset?

This is precisely what O'Shaughnessy (1991, pp. 273-274) asks in his quest for a role for intentions in action. He provides the following example: suppose that on Monday I reach a decision to swim the English Channel on Tuesday at 6 a.m.; I thus form a prior intention that lasts at least until Tuesday at 6 a.m. What then? Does the intention suddenly expire as I start swimming? That seems implausible. Surely the intention is still there, for otherwise I'd be irrational if I was swimming

the Channel without either (a) intending to, or (b) being forced by someone. So, not only the intention is still there as I begin to swim, but, O'Shaughnessy rightly points out, half-way across the Channel I *still have* the intention to swim from Dover to Calais. For, "[d]elete the intention at *any* point [in the course of the action] and the act fizzles out […]" (O'Shaughnessy, 1991, pp. 273-274). That is, at whatever point I ceased to have that intention, I would thereby put an end to my action (e.g., O'Shaughnessy says, "I join my friends in the boat", p. 274). So my prior intention must still be there while the action unfolds.

Now, the example provided by O'Shaughnessy concerns an action, that of swimming the Channel, which is much more complex and extended in time than that of raising one's arm – for one thing, the intention to raise one's arm could be seen as a proper part of that to swim the Channel (depending on the technique employed in swimming). Searle (1991, p. 298) acknowledges that prior intentions can exist and continue to function after the onset of the action, but he adds the proviso "at least in the case of complex extended actions". But why should the case of simple actions such as raising one's arm be different? Supposedly, at any point during my arm raising, I still intend to raise my arm. Thus, after action onset, there is an overlap between prior intentions and intentions in action that puts pressure on the distinction drawn in temporal terms: after action onset, prior intentions seem to become redundant. Wouldn't it be better to posit just one kind of intention, formed prior to the initiation of the action and persisting after action onset? The question now is whether differences in content can preserve the distinction between prior intentions and intentions in action.

2.2 Differences in content

Searle presents the differences in content between the two kinds of intentions in a way that is neutral between the two following very different readings. On the one hand, we may think that the content of a prior intention *needn't be* as detailed as that of an intention in action, and so it normally isn't. On the other hand, a stronger reading of Searle's characterization may be given to the effect that the content of a prior intention *can't be* as detailed as that of an intention in action. It is only the latter reading that provides reasons for thinking that we really have two distinct kinds of intentions there, and I don't think there are grounds for granting such stronger reading.

While Searle does not motivate either reading, Pacherie (2000, 2003) seems to suggest the stronger reading, on the basis of two considerations. First of all, she invokes the cognitive cost that having prior intentions with such a detailed content as the relative intentions in action – a cognitive cost that would, in most cases, ultimately be unnecessary. This point is somehow between the two readings that I proposed: it seems to support the idea that the content of the prior intention both *needn't be* and in general *had better not be* as detailed as that of a prior intention, which by itself doesn't necessarily suggest that it *can't be*.

A second motivation pulls more clearly in the direction of the impossibility for the two contents to be the same, and it has to do with what Pacherie calls the *anchoring to the situation* of the intention, which consists in putting one's intention in touch with the information that becomes available once the corresponding action has started. This process is further divisible into two aspects. First, supposing one has formed the prior intention to get on a bus to go somewhere, when the time comes one must not only get on a certain bus (say, bus n. 12), but on

this bus (this is a kind of indexical content), which will be one of the many passing by the stop where one is waiting. Secondly, more often than not, prior to action onset one is simply not in the position to predict all the possible obstacles that might force one to modify his course of action accordingly (Pacherie, 2003). Now the question is: are these points compelling? In particular, do they suffice to make the content of prior intentions impossible to reconcile with those of intentions in action?

To return to the arm raising example, suppose now that the one who forms the prior intention to raise his arm is Peter Sellers, during the filming of "Dr Strangelove". Here the actor interprets the role of a man (Dr Strangelove) who appears to be affected by the anarchic hand syndrome, a neurological disorder in which the affected person's hand seems to have acquired a mind of its own. Accordingly, in the film we see Peter Sellers's hand every now and then suddenly going up as if it was no longer under his control. Presumably, then, at least the first times he was making this movement during the filming, it is plausible that he could have thought out how to move his hand in such a way as to make it look like that of an anarchic hand syndrome patient. That is, we can imagine that, prior to moving his hand, he must have formed an intention not only to raise his hand, but to do so in a way that could look out of his control – in particular, with a specific dynamics (his hand goes up very quickly). Furthermore, part of the characterization of Dr Strangelove includes that his arm does not just generally move out of his control, but, specifically, moves in such a way as to reproduce the Nazi Party salute. Peter Sellers must have therefore taken care to orient and move his arm accordingly.

In a nutshell, if we reflect on all the kinematic and dynamical characteristics of Peter Sellers's arm moving up, it becomes plausible that, even prior to the

initiation of such movement, he must have been able to form the intention to raise his arm in such a way that the content of his intention was probably as detailed and fine grained as that of his alleged intention in action. In that situation, for the sake of the success of the performance it would be appropriate to work out all the details of the action prior to action performance. Therefore, we wouldn't have a case of useless cognitive cost. Furthermore, between before and after action onset, nothing changes either in terms of indexical content or, presumably, of possible obstacles available – everything is very predictable. This reflection goes to show that there needn't be any difference in content between intentions in action and prior intentions, insofar as the content of the latter, depending on the need, can in principle be as detailed as that of the relative intentions in action. If we do accept this picture of how the story goes, then, the differences in content between prior intentions and intentions in action turn out to be annihilated. But differences in content, together with those in terms of temporal characteristics, were the only distinguishing features of intentions in action as opposed to prior intentions.

Conclusion

Searle (1983) has put forward a distinction between prior intentions and intentions in action to account for both faces of intention, namely future-directed intentions and intentional actions. While I embrace the idea that intentional actions should be explained in terms of intentions, I object to Searle's distinction by proposing that just one kind of intention suffices to do the explanatory job that prior intentions and intentions in action are meant to do. I've argued that these two kinds of intentions are distinguished by Searle on the basis of both temporal and content characteristics. The O'Shaughnessy example showed that, when there

are both a prior intention and an intention in action to do something, the prior intention is still there after action onset, and so it overlaps with the intention in action. Furthermore, on the basis of the Dr Strangelove example, I've pointed to the possibility that the contents of the two intentions may be the same. Hence, there cannot be two different kinds of intentions at play. This suffices to conclude that, whenever there is an action planned in advance, just one kind of intention underlies the whole process—not just at the planning stage, but also at the execution stage. What about spontaneous actions, though? In the light of what has been established in the case of actions planned in advance, it would seem that the best explanation for spontaneous actions lies in the same kind of intentions as those underlying action planned in advance—only, the intentions underlying spontaneous actions are shorter-lived.

§6. What are intentions-with-which? The hierarchical structure underlying intentions

Introduction

In the course of the previous chapters, motor intentions have been introduced as standing in a specific relation to motor representations—either referring to action outcomes by deferring to action outcomes, or as having the same content as a motor representation—as well as providing the grounds for arguing for the continuity between prior intentions and intentions in action. In this chapter, motor intentions will be seen to occupy the lowest levels of a hierarchical structure underlying intentions. This will give me the chance to explore the notion of *intention-with-which*, and to review a series of experimental results in the light of the introduced hierarchical structure.

1. Three notions of *intention-with-which*

As a matter of fact, one of the early versions of the leading theory in philosophy of action, the causal theory of action (Davidson, 1963), had it precisely that intentions with which are simply ways of redescribing what someone is doing in terms of a *primary reason*, conceived as a pro-attitude towards actions having some feature *F*, together with the belief that the action in question has that feature. It is by virtue of a relation to a primary reason that actions count as intentional. Hence, a reductive theory was given of intentions-with-which and of intentional actions to a theory of action featuring primary reasons (see chapter 2, section 2).

It is possible to distinguish among different senses of intention-with-which, following Bratman (1987, pp. 128–ff.) on the notion of acting with an intention[45]. Bratman first traces a distinction between what he calls *acting with an intention* and *acting with a further intention*. An example for the first case is opening the curtains with the intention of getting more light (Bratman, 1987, p. 128). In this first example, the act of opening the curtains just *is* an act of getting more light, given the circumstances. Note that, therefore, opening the curtains and getting more light roughly occupy the same time interval. Also, opening the curtains is one of the possible means to getting more light, and is one of the potentially many ways in which this end could be achieved (for instance, one may have switched the lights on instead). Turning now to acting with a further intention, an example of this is given by opening the curtains with the intention of washing the windows (Bratman, 1987, p. 128). In this second example, the two acts—opening the curtains and washing the windows—occupy distinct time intervals, such that the time interval occupied by washing the windows follows (while presumably not overlapping with) that of opening the curtains. Note that washing the windows could in principle be characterized as an overarching action made of many steps, one of which is opening the curtains, in which case it wouldn't be true that the two acts occupy subsequent time intervals, but rather nested time intervals (that corresponding to opening the curtains included into that corresponding to washing the windows). For my earlier description of the time intervals to work, the act of washing the windows in this specific example has to be interpreted as less inclusive, and specifically as beginning strictly after the time interval corresponding to opening the curtains.

[45] These distinctions were sketched in chapter 2, and are now going to be developed in greater detail.

Once the distinction between acting with an intention and acting with a further intention is in place, Bratman proceeds to distinguish a strong and weak reading of the former (pp. 129-ff.). According to the strong sense of acting with an intention, in the example given beforehand one has the intention of getting more light. Acting with an intention in the weak sense, by contrast, implies that one strictly speaking does not intend to get more light, but rather merely endeavours to get more light. As Bratman (1987, p. 130) rightly highlights, the contrast between the strong and weak sense of acting with an intention is not merely verbal, but rather points to a difference in the webs of norms and regularities to which intending and endeavouring are subjected. I've talked about these differences in chapter 2, so I'm not going to dwell on them again. Following Bratman (1987, p. 133) I'm simply going to stress that in typical cases one endeavours to *A and* intends to *A*, and, for this reason, in what follows I'm mainly going to be concerned with the strong sense of acting with an intention.

Let me now notice that, by distinguishing among acting with a further intention and acting with an intention in the strong sense, it is possible to order actions on the basis of two corresponding principles. As far as acting with a further intention is concerned, as mentioned beforehand if one *A*s with the further intention to *B*, then *A* and *B* will occupy subsequent time intervals, with *A* preceding *B*. So we have a binary relation *further-intending-to(x,y)* where *x* and *y* are actions such that *x* and *y* occupy distinct time intervals I_x and I_y such that I_x precedes I_y. Pairs of actions satisfying *further-intending-to(x,y)* can thus be ordered along a horizontal axis where the variable is time. What about the principle for ordering actions based on acting with an intention in the strong sense? This is more complicated and will occupy the following sections.

2. Hierarchically structured actions

The notion of action hierarchy is widespread in psychology, cognitive science and neuroscience (Shallice, 1988; Csibra, 2007; Hamilton & Grafton, 2007), and is based on the idea that actions can be structured hierarchically in terms of goals and sub-goals. This notion has been referred to on several occasions throughout the thesis, so I'm not going to give further examples of it at this stage. It is now time to make the notion of action hierarchy more precise. In what follows, I'm going to rely on Goldman (1970) insofar as he's provided an especially clear and precise way to render the notion of action hierarchy under the philosophical viewpoint.

2.1 "A fine-grained procedure for individuating acts"

Let me start by a methodological caveat: Goldman's theory relies on a fine-grained individuation of actions. I'm going to describe this interpretation and adopt it for the sake of argument without actually embracing it: as I'm ultimately interested in arriving at a hierarchical structure for intentions, nothing of what I'm going to propose actually hinges on whether the fine-grained individuation of action holds.

Goldman (1970, p. 1) proposes the following example. John at the same time (i) moves his hand, (ii) scares a passing by fly, (iii) moves his hand, (iv) moves his queen to a certain slot, (v) checkmates his opponent, (vi) gives the said opponent a heart attack, (vii) wins his first game of chess. As Goldman points out, theorists such as Davidson and Anscombe have claimed that in this example (i)-(vii) do not point to different act performed by John, but rather to only one act of his, of which (i)-(vii) constitute different descriptions. This is what Goldman terms the *identity thesis*, according to which whenever someone could in principle be said

to be doing many different things at the same time, as in the above example, as a matter of fact she's doing just one thing, of which different descriptions may be given (Davidson, 1963, p. 686, quoted in Goldman, 1970, p. 2).

Goldman argues that the identity thesis faces a series of insurmountable difficulties. The first is due to the fact that two identical acts, by the principle of the indiscernibility of identicals, should exemplify the same properties, and it is possible to find acts that should be identical according to the identity thesis and which at the same time fail to exemplify the same properties. I'm not going to dwell on the examples taken to establish this point (see Goldman, 1970, pp. 3-4), as the latter is not crucial for my purposes. Instead, I'm going to dwell on another reason why the identity thesis is thought not to hold by Goldman. He reasonably claims that, if two acts *A* and *A'* are identical, it cannot be the case that one bears an asymmetric and irreflexive relation to the other. For suppose there exists a binary relation *R(x,y)* such that it is the case that *R(A,A')*, and at the same time it is not the case that *R(A',A)*, or that *R(A,A)*, or that *R(A',A')*. If it is true that *A* = *A'*, then, by substituting *A* with *A'* and *A'* with *A* in *R(A,A')* one obtains precisely *R(A',A)* (and similarly for *R(A,A)* and *R(A',A')*). But an asymmetric and irreflexive relation holding between allegedly identical (if we rely on the identity thesis) acts can indeed be found according to Goldman, and it is of special interest to us.

According to the identity thesis, in the example given at the beginning of this section checkmating one's opponent and moving the queen to king-knight-seven are identical acts. We may well say that in the situation described by the example one checkmates his opponent *by* moving the queen to king-knight-seven. So, the relation *by(x,y)*, defined as *x* by *y*, holds of (checkmating one's opponent, moving the queen to king-knight-seven). Note that this relation is asymmetric, as it

is not the case that one moves the queen to king-knight-seven *by* checkmating one's opponent. This relation is also irreflexive insofar as it is not the case that one checkmates one's opponent by checkmating one's opponent (and similarly for moving the queen to king-knight-seven). Since an asymmetric, irreflexive relation is found to hold of supposedly identical acts, and since asymmetric, irreflexive relations cannot hold of identical acts, according to Goldman we should deny that acts such as checkmating one's opponent and moving the queen to king-knight-seven in the above given example are identical acts. This piece of reasoning leads Goldman to adopting what he calls a *fine-grained procedure for individuating acts*, such that, for instance, checkmating one's opponent and moving the queen to king-knight-seven are *different* acts. As mentioned at the outset, I'm not going to be concerned with defending the fine-grained procedure for individuating acts, but I'll limit myself to assuming it for the sake of argument. This will allow me to follow Goldman in the deepening of the relation that can be said to hold between pairs of acts such as checkmating one's opponent and moving the queen to king-knight-seven—namely what Goldman calls *level-generation*, which, as I'm going to show, with appropriate modifications can help us elucidate the notion of acting with an intention.

2.2 What is level-generation? Is there a more useful notion for structuring actions?

Level-generation (Goldman, 1970, pp. 20-ff.) is an asymmetric, irreflexive and transitive relation. A definition of this notion is going to take most of this section, and the subdivision of such notion into four categories. After that, I'll propose my own revision to that notion. Let us start by noting the following. We've

already seen that checkmating one's opponent and moving the queen to king-knight-seven in the previous example are related in such a way that it is appropriate to say that one checkmates one's opponent *by* moving the queen to king-knight-seven. In general, according to Goldman (1970, p. 21), whenever the *by* relation holds between two actions, level-generation between those can be said to obtain. This is reasonable insofar as, as we've seen, the *by* relation is asymmetric and irreflexive. It can also be said that the *by* relation is transitive: e.g., once one checkmates his opponent by moving the queen to king-knight-seven and moves the queen to king-knight-seven by moving one's hand, it can also be said that one checkmates his opponent by moving one's hand[46]. But cases in which the by relation hold do not exhaust the totality of the cases in which level-generation obtains. So let us see what further characterizes this relation.

Goldman (1970, pp. 21-22) specifies two temporal properties of acts that are related by level-generation. First of all, whenever two acts are related by level-generation, they are always done at the same time: none is subsequent to the other. This is the case for the aforementioned example of checkmating one's opponent by moving the queen to king-knight-seven. It is not the case that one moves the queen to king-knight-seven *and then* checkmates one's opponent (or conversely). The same goes for Bratman's example concerning acting with an intention: it is not the case that one opens the curtains *and then* gets more light.

[46] I'd like to point out that the transitivity of the *by* relation is not wholly unproblematic. For instance, while it is strictly speaking true that in the given example one checkmates one's opponent by moving one's hand, it is also the case that checkmating one's opponent by moving the queen to king-knight-seven seems to be much more informative than checkmating one's opponent by moving one's hand, and it seems that one may not accept: "By moving his hand" as an answer to the question: "How did he checkmate his opponent?", on the grounds that the former is not sufficiently informative to count as an appropriate answer (any chess move, and not just checkmating, is going to involve moving one's hand).

Rather, the two things are done simultaneously by the same individual. Secondly, working on the assumption that the two actions related by level-generation are done at the same time, it is not the case that the two actions in question are independent. That is, it shouldn't be possible to say of two level-generational acts that one is performed *while also* carrying out the other. In Goldman's words, the two actions related by level-generation shouldn't be *co-temporal*. Indeed, it is incorrect to say that one moves the queen to king-knight-seven *while also* checkmating one's opponent, or that one opens the curtains *while also* getting more light. This is because, as mentioned at the outset, given the circumstances opening the curtains (e.g.) just *is* getting more light.

So far, these two temporal characterizations ensure that any two level-generational acts are going to occupy the same interval of time. But now we need to be told more about the nature of level-generation. In order to do so, we need to separately look at four categories of level-generation that Goldman considers to be exhaustive of all cases of level-generation. As I proceed to illustrate the differences among them, the nature of level-generation is going to become clearer. I'll also mention my reservations about certain aspects of the taxonomy proposed by Goldman, and this will give me the opportunity to introduce the details that are still missing in Goldman's framework to obtain the hierarchical structure that, I'm going to propose, underlies intentions.

Goldman (1970, pp. 22-ff.) distinguishes among four kinds of level-generation: (1) causal generation, (2) conventional generation, (3) simple generation, (4) augmentation generation. In the case of causal generation, the generated action is always one that can be described in terms of *causing E* or *bringing about E*, where *E* is an event or state of affairs (Goldman, 1970, p. 23). For

instance, in a given situation in which I flip the switch and thereby turn on the light, my flipping the switch causally generates my turning on the light, where the latter act can be rendered as *causing the light to go on*. Causation is involved insofar as one's causally generating act *A* causes a certain event *E* (my flipping the switch causes the lights to go on), and causally generates a certain act *A'* (my flipping the switch causally generates my turning on the light), and one's causally generated act *A'* consists in one's causing the event *E* (Goldman, 1970, p. 23). Still, as Goldman warns, one's causally generating act does *not* cause one's causally generated act (my flipping the switch doesn't *cause* my turning on the light—in the given circumstances, it *is* my turning on the light). More will be said about the contrast between causation and causal generation in due course.

Conventional generation (Goldman, 1970, p. 25) relies on social conventions or practices making it the case that, in appropriate circumstances *C*, act *A* counts as act *A'*. For instance, extending one's arm out of the window while driving counts as signalling for a turn. According to Goldman (1970, p. 26), this relies on the presence of appropriate circumstances *C*—driving a car as opposed to standing in one's garage with one's car engine switched off and one's car sitting still—such that, in those circumstances, a rule *R* makes it the case that extending one's arm out of the window while driving counts as signalling for a turn. As for the rules underlying cases of conventional generation, Goldman subdivides them into normative and non-normative. Normative rules have to do with specifying acts that are obligatory or forbidden, and give rise to examples such as one in which *S*'s trying to save his friend's life conventionally generates *S*'s doing his duty. Non-normative rules rest on institutional frameworks or games dictating the significance of certain acts in certain circumstances, and are at the basis of the turn

signalling example, or of the example involving checkmating one's opponent by moving one's queen to king-knight-seven. As we'll see, cases of the latter sort, in comparison with those involving normative rules, are easier to accommodate in the framework that I'm going to propose.

Simple generation (Goldman, 1970, p. 26), by contrast with causal generation, doesn't involve causation at all: it relies solely on the presence of circumstances that make the performance of an act *A* also the performance of an act *A'*. For instance, in circumstances in which George has jumped six feet, my act of jumping 6 feet 3 inches counts as outjumping George (Goldman, 1970, p. 26). Goldman contrasts simple generation not only with causal generation, but also with conventional generation, since simple generation doesn't appeal to the existence of rules. Still, to the extent that simple generation relies on the presence of the appropriate circumstance, I would say that there is some degree of similarity between simple and conventional generation.

Let us now turn to agumentation generation (Goldman, 1970, pp. 28-ff.), which is where my main disagreement with Goldman lies. Here are a few of the examples of augmentation generation that he provides. *S*'s running at 8 m.p.h. is level-generated by augmentation by *S*'s running. *S*'s jump-shooting is level-generated by augmentation by *S*'s shooting (in a basketball context). Now, Goldman claims that the generated act is obtained by, so to speak, *augmenting* the generating act with a certain manner or with certain circumstances in which it is performed. Goldman claims that this form of level generation is significantly different from the other three, insofar as, for instance, the *by* relation doesn't capture it—by his lights, we would not normally say that S ran at 8 m.p.h. *by* running, or that S jump-shot *by* shooting. And herein lies the mistake, I claim. For,

instead, I think we can agree that it would be admissible to say that S ran *by* running at 8 m.p.h., or, even more clearly, that one shot *by* jump-shooting, in the following sense. Running at 8 m.p.h. is one of the possible ways to run. Likewise, jump-shooting is one of the possible ways in which one may shoot. Goldman's characterization of augmentation generation is not sensitive to this feature, whereas one thing that seems to be implicit in his notion of level-generation more generally is that to which I've made reference ever since the first chapters, namely a means-end relation. To be precise, cases of augmentation generation are going to be more appropriately described in terms of determinate-determinable relation, in the following sense. Cases of what Goldman terms of augmentation generation are such that, e.g., one cannot run at 8 m.p.h. without running. Running at 8 m.p.h. is thus best described as a *further determination* of running, as opposed to a *means* towards running. The relevant similarity between means-end relations and determinate-determinable relations is that for every couple of acts *(x,y)* satisfying either kind of relation, *x* is always going to be more determinate and constrained than *y*. In what follows, I'm going to use the term *means-end-generation*, with the understanding that, whenever what Goldman terms augmentation-generation is involved, means-end-generation has to be interpreted as determinate-determinable-generation. With these ideas in place, if we look back at all the other cases of level-generation, it seems that we can analyse all of them in terms of what I'm going to term means-end-generation.

Take causal generation. Opening the curtains is one of the possible means to getting more light (alternatively, one could have switched on the light). As for conventional generation, cases of non-normative generation are easier to accommodate in terms of means-end relations. In the car example, extending one's

arm out of the window is one of the possible ways of signalling for a turn—another could be resorting to the indicators. It is not so straightforward to establish the idea that, for instance, saving a life is one of the many ways of doing one's duty. It seems that in the given example one is not simply seeking for a way to do one's duty, which may likewise be fulfilled by throwing an envelope in the appropriate bin as opposed to throwing it on the ground, but rather one wants to do one's duty with respect to the threatened life. But, as it'll become clear in the course of this chapter, it's much more concrete cases, which happen to be captured in non-normative vs. normative rules, that I'm interested in, so I think we can live with the difficulty that normative rules seem to pose. As I said earlier on, augmentation generation, by contrast, had better be characterized in terms of determinate-determinable-generation, with the outcome that the order of generation among acts linked by augmentation gets reversed with respect to Goldman's framework.

Curiously enough, Goldman considers the option that I'm proposing, namely that augmentation generation should be conceived as operating in the opposite direction, but rejects it on the following grounds.

> In all of the other species of generation, a generated act is formed by making use of some *additional* fact not implicit in the generating act. This feature of generation is preserved by saying that S's running generates S's running at 8 m.p.h.; it would run contrary to this feature to say that S's running at 8 m.p.h. generates S's running.
> (Goldman, 1970, p. 29)

Still, there are important reasons in favour of my interpretation. First of all, all other three forms of level-generation can be re-read in terms of a means-end /

determinate-determinable structure, as I've suggested. In addition to that, as Goldman himself notes (1970, p. 30), the four species of level-generation are intended to be exhaustive and not mutually exclusive: when two acts are related by level-generation, it doesn't need to be the case that one and only one species of level-generation is involved. For instance, according to Goldman (1970, p. 30) one's giving her opponent a heart attack by moving one's hand involves causal generation, conventional generation and augmentation generation (presumably by means of all the intermediate generated levels). More to the point, again by Goldman's own admission, many cases of generation straddle the boundary across two or more categories of level-generation. For that to be possible, the four categories must be sufficiently similar, and Goldman's interpretation, despite what he claims, would run counter this similarity. On the contrary, interpreting augmentation generation according to my suggestion would lead all four categories of level-generation to share a means-end / determinate-determinable structure, thus enabling instances of level-generation to fall on a borderline across different categories.

3. Are actions fulfilling motor intentions the same as basic actions?

On the basis of the preceding section, we're now ready to describe a hierarchical structure in which actions are related by means-end-generation of various kinds. To go back to our initial chess example, the actions involved in that situation can be ordered by means of the following structure (see Goldman, 1970, p. 31):

(i) Giving one's opponent a heart attack

(ii) Checkmating one's opponent

(iii) Moving one's queen to king-knight-seven

(iv) Moving one's queen

(v) Moving one's hand

In this structure, each action (n) means-end-generates action (n-1). One may wonder whether this sort of structure constructed by means of means-end-generation ever bottoms out, and, if so, where. That is, could we once more ask how (v) is performed and thus obtain a sixth action in reply? Could we then re-apply this strategy over and over? Until when can we do so and obtain as an answer an event that a subject can voluntarily initiate (thus excluding answers along the lines of producing neuron firings in one's brain[47])?

In order to answer these questions, it is now time to tackle the notion of *basic action*. We've already explored at length the idea that there are actions that we perform *by* performing other actions—e.g., checkmating one's opponent *by* moving one's queen to king-knight-seven. The notion of basic action, roughly, has to do with the idea that there are actions that are *not* performed by means of other actions (see Goldman, 1970, p. 6). A methodological caveat: the following discussion will rely on the idea that we have at least an intuition, along the above mentioned lines, as to what a basic action may be. This intuition is going to guide our quest for a definition of basic action, a definition that will enable us to include what we think should count as a basic action and to rule out what should not count as a basic action, either because it's too complex and/or abstract for being generated without performing a more basic action (e.g., becoming a good person is not a basic action), or because we cannot consciously initiate it (e.g., producing

[47] This example is due to Searle (1983, p. 99).

neuron firings in one's brain is not a basic action—it's not an action at all—and contracting certain muscles in one's arm and hand[48] for most people are not basic actions[49]).

The notion of *basic action* was introduced by Danto (Danto & Morgenbesser, 1963; Danto, 1965, reported in Goldman, 1970, p. 24), and defined in the following way:

> *A* is a basic action if and only if
>
> (1) *A* is an action, and
>
> (2) whenever *S* performs *A*, there is no other action *A'* performed by *S* such that *A* is caused by *A'*.

Goldman rightly notes that if the notion of basic action is meant to do justice to the idea that there are actions that are not performed by means of other actions, then Danto's formulation, as it stands, is inadequate. Goldman claims that the mistake lies in appealing to causation instead of causal generation. Now, as pointed out beforehand, the notions of causation and causal generation are indeed related, but they're also importantly different. Essentially, whenever causation between two events is involved, this normally implies that the causing event precedes the caused event. On the basis of the temporal characteristics of level-generation, this means that two events related by causation cannot be related by causal generation.

Goldman (1970, p. 24) points out that appealing to causation instead of causal generation results in failing to classify as basic actions acts that we'd like to

[48] Another example due to Searle (1983, p. 99).

[49] But see Pacherie (2008, p. 199).

so classify, and conversely. For instance, suppose that I have a gun in my hands and that, all of a sudden, I kill someone without there being anything that caused me to do so (this person hasn't either frightened me, or provoked me in any way, nor has there been anyone who forced me to act this way). On Danto's definition of basic action, my killing this person is a basic action. And yet, we can draw a diagram analogous to the one concerning the chess example, which will reveal many layers below killing this person: I've killed this person by firing a gun, and, in turn, I've fired my gun by pulling the trigger, I've pulled the trigger by contracting my fingers in a certain way, and so on. So there are many actions by means of which I kill this person, and on the basis of this we wouldn't want to classify killing a person as a basic action. Note that killing a person, firing a gun, pulling the trigger and contracting my fingers are related by causal generation, and not by causation. As for the consequence that we are prevented from classifying as basic actions acts that we'd like to so classify, suppose for the sake of argument that the earlier mentioned hierarchy terminates with contracting my fingers. Suppose also that, by contrast with what was said about the earlier example, we're in a different situation in which I've agreed I'm going to fire my gun as soon as an accomplice of mine orders me to do so. In this modified scenario, then, the contraction of my fingers is caused by my accomplice's order[50]. So, on Danto's definition, we cannot classify the act of contracting my fingers as a basic action, although we would like to do so.

Goldman (1970, pp. 63–ff.) proposes a revised notion (with respect to Danto's formulation) of basic action that relies essentially on two features: being able to generate an action *A* from one's *want* to carry out this action, and the

[50] This is a variation on an example given by Goldman (1970, p. 74) to establish the same conclusion.

independence of this generation from level-generational knowledge, or cause-and-effect knowledge, except when it comes to knowing that one's wanting to generate *A* is going to cause one's *A*-ing[51]. Let us see what justifies these two requirements. First, note that Goldman takes the term *want* to be

> roughly equivalent to "feeling favorably toward *x*," "being inclined toward *x*," "being pro *x*," "finding *x* an attractive possibility," "finding *x* to be a 'fitting' or 'appropriate' possibility," etc.[52]
>
> (Goldman, 1970, p. 49)

Accordingly, Goldman's reference to a *want* as the origin of a given action ensures that the agent has a positive attitude towards the action in question: it is not an involuntary action, but rather one that the subject wishes to carry out. As for level-generational knowledge and cause-and-effect knowledge, Goldman observes that, whenever we mean to carry out an action *A'* that is not basic, we need to *want* to perform an act *A* that is either going to cause *A'* or level-generate *A'*. So, for instance, if our goal is to pull faces we'll need to contract our facial muscles in a

---

[51] Here is Goldman's definition (see Goldman, 1970, pp. 10-ff. on act-types and act-tokens):

> *Property* A *is a basic act-type for* S *at* t *if and only if:*
>
> (a) *If* S *were in standard conditions with respect to* A *at* t*, then if* S *wanted to exemplify* A *at* t*,* S*'s exemplifying* A *at* t *would result from this want; and*
>
> (b) *the fact expressed by* (a) *does not depend on* S*'s level-generational knowledge nor on* S*'s cause-and-effect knowledge, except possibly the knowledge that his exemplifying* A *would be caused by his want.*
>
> (Goldman, 1970, p. 67)

[52] Goldman subsequently (1970, pp. 86-ff.) gives a detailed treatment of the notion of *wanting*, but I leave it aside as, for reasons to be explained shortly, it is best to substitute this notion with that of *intending*, in line with the theory developed throughout my thesis.

certain way, stick our tongue out, and so on, and all these level-generational acts will need to be voluntarily initiated. Wanting to pull faces, in short, results in initiating a series of acts that are level-generationally more basic, and which in turn require voluntary initiation. By contrast, Goldman claims (1970, pp. 65-66), if we want to move our hand, as far as our voluntary initiation goes there's nothing level-generationally more basic that we may want to do. Analogously, there are acts such as vomiting (Goldman, 1970, p. 66) that cannot be induced at will without voluntarily initiating acts that are causally more basic, such as putting one's fingers down the throat. Acts such as vomiting do not therefore count as basic actions.

Now, I acknowledge the improvement of Goldman's notion of basic action over Danto's as for capturing the idea of an action that is not performed by means of any other action, but I think that further modifications to Goldman's formulation are possible, which will make the notion suited to my purposes. First of all, when it comes to the *want* requirement, we're now equipped with a sufficiently powerful notion of intention to make that requirement even more efficient in capturing the notion of basic action. So we're going to have as a first requirement that for an action *A* to be a basic action for a subject *S*, *S* has to be able to *intend* to perform *A*, and performing *A* is going to follow from one's intention to *A*. For all that has been said throughout the present work, appeal to intentions has a number of implications. First of all, an intention is much more than a favourable attitude towards an action, insofar as, for instance, intention defies reconsideration, whereas a mere favourable attitude not necessarily does (and most likely won't). Also, we have a much better grasp on how performing an action *A* follows from the corresponding intention.

Most importantly for present purposes, we know that having an intention implies being able to introspect the content of that intention, so we can only have intentions whose content we can introspect. This feature of intentions leads us to put a lower boundary on actions that qualify as candidates for being basic, in the following way. Suppose I execute a certain hand movement—say, while playing the piano. That hand movement can be characterized at one level (e.g.) as an alternating contracting of my index and middle finger. At a means-end-generational level below, it can be characterized as an alternating contracting of my index and middle finger at a certain angle and at a certain speed. Should our expressive capacities come to an end, we may still in principle resort to demonstrative concepts. Characterizations at means-end-generational levels further below are going to be possible depending on our degree of awareness of our actions, on our introspective abilities, which are going to be different across individuals. The introspectability requirement on intentions is going to ensure that, in accordance with individual variability in introspective abilities and awareness, it'll be possible to characterize as basic actions only events that a given subject is capable of voluntarily initiating. Some events will be beyond the capability of voluntary initiation for most individuals, whereas other events might present an interesting variation across individuals. On the latter point, Pacherie (2008, p. 199) notes that contracting certain muscles cannot be a basic action if not (perhaps) for some yogi. Interestingly, she notes, for most people it won't be a basic action insofar as one may need to voluntarily initiate other actions, such as raising one's arm, in order to obtain the contraction of the muscles in question (either by means-end-generation or by causation, I'd like to add).

As for Goldman's second requirement concerning the independence of an action from level-generational or causal knowledge, in the light of my earlier considerations about level-generation I'm going to substitute the latter with means-end-generation. So I'm going to suggest as a second requirement for an action A to be a basic action for subject S that there mustn't be any action A' means-end-generationally more basic than A that the agent may intend to perform. Prima facie, my proposal looks remarkably similar to Searle's (1983) characterization of basic actions according to which:

> *A* is a basic action type for an agent *S* iff *S* is able to perform acts of type *A* and *S* can intend to do an act of type *A* without intending to do any other actions by means of which he intends to do *A*.
> (Searle, 1983, p. 100)

Still, there are a few relevant differences. As for Searle's second requirement, which would seem to be identical with mine, it's worth recalling that the way in which I've developed the notion of intention significantly diverges from Searle's[53].

One thing that, according to Searle (1983, p. 100), follows from his definition is that it makes the notion of basic action dependent on the agent and on her skills, in such a way that an action which is basic for a given agent may not be likewise basic for another agent. Searle provides the following example: for a good skier, making a left turn can be a basic action, insofar as he no longer needs to worry about (I take it, in the sense of voluntarily initiate while consciously attending to) the relevant means-end-generational actions, such as shifting weight

[53] See chapter 5 for my rejection of Searle's (1983) distinction between prior intentions and intentions in action.

from one ski to another, and so forth. The beginner, by contrast, in order to make a left turn will need to consciously attend to all these features, being unable to bypass them and exclusively focus on the means-end-generational levels above[54]. Does the same result follow from my definition? Not quite, and I'll explain why.

Recall the discussion that I made in chapter 3 about the accessibility of a mental state to introspection. There I insisted on the idea that accessibility to introspection of the content of a mental state at a given point—say, during action execution—does not have any bearing on the question as to whether that mental state is an intention. A certain mental state may well be in intention, while its availability to introspection is limited on certain occasions due to the demands related to a successful performance (see the notion of *flow* introduced in chapter 3, section 3.3). The take-home message relevant to the present circumstances is that accessibility of a mental state to introspection may vary from time to time depending on the circumstances. Skill, however, is a different variable with respect to capability for introspection. While it is true that one often ends up introspecting at a less deep level once she's become more skilled, becoming more skilled doesn't straightforwardly imply a change in the basic action that one is able to perform, unless acquiring a skill is accompanied by a flat-out and irredeemable change in capabilities for introspection. That's why Searle's point about basic actions varying with the subject's skill needs to be handled with more care.

Let us now take stock. On the basis of Goldman (1970), with a series of modifications, I've introduced the idea that, for any action that one may be said to perform at a given time, one is simultaneously performing a series of other actions that are ordered in a hierarchy on the basis of the means-end-generation principle.

[54] Pacherie (2008, p. 199) provides an example in the same vein, according to which executing a trill may be a basic action for a skilled pianist, but not for the novice.

By means of the notion of basic action, I've suggested that each hierarchy bottoms out. Now recall some examples of action hierarchies that I gave earlier on in the chapter, and recall the lowest action that I described in each of them—which is not necessarily where they ended, as they may in principle have ended further down. The hierarchy corresponding to the chess example was specified down to moving one's hand. In the gun firing example, the lowest specified level involved contracting my fingers. These levels all seem to involve actions which, insofar as they essentially involve a movement of one's body, are apt to fulfil motor intentions—let's call them *motor acts*. Even leaving it open whether the described actions are basic or not, actions that are even more means-end-generationally basic are most likely going to involve motor acts. Should we then conclude that basic actions just are motor acts?

Things are far from being so straightforward, but there certainly is an interesting intersection between basic actions and motor acts. Even supposing for the sake of argument that whenever an action is basic it is a motor act, taking into account the variability across individuals discussed at length has the consequence that, with respect to a given action type, for different people, different actions, that are both motor actions, may be basic. For what I've said so far, it could at best be the case that basic actions are a non-trivial subset of motor acts.

4. Structuring intentions on the basis of actions

Let us now take stock. We started with the observation that there are at least two senses of acting with an intention, namely *acting with an intention* (which in turn can be interpreted in a strong and in a weak sense) and *acting with a further intention*. We have seen that it is possible to order actions along two axes following

these two principles. Acting with a further intention provides an ordering of actions along a horizontal axis, such that *A*ing with the further intention of *B*ing implies that action *A* precedes action *B*—if one opens the curtains with the further intention of washing the windows, opening the curtains precedes washing the windows. By contrast, acting with an intention tout-court provides an ordering of actions along a vertical axis, which I've been calling an action hierarchy structured on the basis of means-end-generation. Whenever an action *A* is a means towards accomplishing action *B*, *A* can be located below *B*, and this vertical ordering comes to an end with a basic action.

Now, so far I've followed Goldman's idea that actions related by means-end-generation occupy the same interval of time. Things may not necessarily be that simple, however, if we consider that a single action may be means-end-generated by a series of more than one action. Recall that, at the beginning of this chapter, I suggested an alternative interpretation of Bratman's example involving opening the curtains in order to wash the windows. According to that interpretation, washing the windows could in principle be characterized as an overarching action made of many steps, one of which is opening the curtains, the latter therefore occupying an interval of time which is strictly less extended than that corresponding to washing the windows. Working on these assumptions, the action of washing the windows can be thought to be means-end-generated by a series of actions less extended in time, such as opening the curtains, dipping a cloth in a bucket of water, and so on.

Now, there are deep and difficult questions as to how actions should be segmented along a given level (e.g., should there be an action of reaching for the cloth between opening the curtains and dipping the cloth in water? Or does

reaching for the cloth belong to a level means-end-generationally below dipping the cloth in water?) Experimental evidence shows that the subdivision of an event such as an action into meaningful steps is routinely performed by human beings, as witnessed by research on behaviour parsing (introduced by Newtson, 1973). Research on how individuals segment the actions they observe into more basic chunks has yielded controversial results (see Kurby & Zacks, 2008). The question as to how actions are to be segmented both as far as their production and their observation are concerned still needs a lot of work. For the time being, it is plausible to suppose that (i) the question as to which action units one is able to produce has a bearing on that as to which action units one is able to perceive, and conversely (many results concerning mirror neurons and mechanisms can be interpreted along these lines—see Cattaneo et al., 2007), and (ii) partly as a consequence of (i), it could be the case that action segmentations are going to vary across individuals, or even across a single individual's lifetime.

In spite of all the difficulties that there might in practice be in pinpointing the correct segmentation for a given action performed by a given individual, the proposal of a hierarchical structure for action that extends both along a horizontal and a vertical dimension should by now have proved coherent. I now wish to propose a hierarchical structure underlying intentions that is going to mirror the hierarchical structure underlying the corresponding actions. First of all, it is possible to provide different descriptions for an action along a vertical dimension, where any two adjacent levels are related by level generation. Basic actions have been defined as the foundations of this hierarchy, and motor acts have been characterised as candidates either for basic actions or for actions level generated by basic actions, where the latter are voluntary movements. Secondly, an action,

such as reaching to grasp a cup of coffee for drinking, may be segmented along a horizontal dimension, for example into reaching towards a cup, grasping it, bringing it to the mouth, and drinking from it. This has provided the context for introducing motor actions as series of motor acts organised in a chain. Now let's see what may be gained under the explanatory viewpoint by adopting the distinctions introduced.

5. Re-interpreting experimental results in the light of a hierarchical structure for intentions

5.1 Re-interpreting results by resorting to the vertical dimension only

I'm going to start by reviewing some experimental results as well as some hypothetical scenarios that can be interpreted by means of what I've called the vertical dimension. I'm going to show that the experimental results as well as the hypothetical scenarios in question can be read as involving a layering of intentions—sometimes just motor, sometimes both motor and non-motor.

Mirror neurons have been recorded (Gallese et al., 1996) that represent not only grasping tout-court, that is, with whatever effector, but also some that respond only to grasping with a specific effector, such as one's hand (regardless of which hand is being employed), and some that fire only in the case of *grasping with a precision grip* or *grasping with a whole-hand prehension*. Those that fire in correlation with grasping with the hand, independently of the kind of grip, have been called *broadly congruent mirror neurons*, by contrast with those that fire in correlation with a specific kind of grip, which have been termed *strictly congruent mirror neurons*. Notice that the names that have been chosen for these kinds of mirror neurons reflect an implicit understanding that *grasping with the hand* is the

outcome, whereas *precision grip* and *whole-hand prehension* are possible means to achieve that outcome. Likewise, *grasping* may be seen as an outcome for which *grasping with the hand* and *grasping with the mouth* are possible means to achieve that outcome. Thus, the above reported actions can be ordered hierarchically as follows:

(A1) grasping;
(A2) grasping with the hand (or grasping with the mouth);
(A3) grasping with the hand using a precision grip (or grasping with the hand using a whole-hand prehension).

A grasping with the hand using a precision grip, therefore, can be described at least at the following three different levels. These are ordered hierarchically in such a way that grasping, at level (A1), is means-end generated by either grasping with the hand or grasping with the mouth, at level (A2), and grasping with the hand at (A2) is means-end generated by either grasping with the hand using a precision grip or by grasping with the hand using a whole-hand prehension, at level (A3). Notice that the labels I've employed, (A1)-(A3), are not meant to imply that the corresponding levels are adjacent. I.e., I'm leaving it open that there could be one or more action descriptions that are means-end-generationally intermediate between, say, grasping with the hand (level (A1)) and grasping with the hand either with a precision grip or with a whole-hand prehension (level (A2)). Means-end generation applies insofar as (A1), (A2) and (A3) occupy the same time interval, and since means-end-generation is a transitive relation, even if there were further intermediate level between (A1) and (A2) or between (A2) and (A3), this

wouldn't lead to the denial of (A1), (A2) and (A3) being related by means-end-generation. It is important to note that (A1), (A2) and (A3) are related by that species of means-end-generation that involves determination as opposed to a proper means-end relation: it is not possible, e.g., to grasp with one's hand by using a precision grip without grasping with one's hand. The former action description is more determinate than the latter, rather than, strictly speaking, constituting a means towards its realization.

In the light of the earlier considerations, the following intentions may be specified as underlying (A1), (A2) and (A3), respectively:

(I1) the intention to grasp;
(I2) the intention to grasp with the hand, or the intention to grasp with the mouth;
(I3) the intention to grasp with the hand using a precision grip, or the intention to grasp with the hand using a whole-hand prehension.

Recall the questions that were posed in the first chapter (section 4): for any given action, how many intentions are being fulfilled? And what kind of intentions? Take an instance of an action of grasping with the hand using a precision grip. This action fulfils at least three intentions, namely: the intention to grasp (level (I1)), the intention to grasp with the hand (level (I2)), and the intention to grasp with the hand using a precision grip (level (I3)). Given their content, which involves sequences of bodily configurations, these are all motor intentions. The same proviso that I made for action levels (A1)-(A3) holds for (I1)-(I3): no assumption is being made about the existence of intentions at intermediate levels, so, at least in principle, there may well be more (motor) intentions than an act of grasping with

the hand using a precision grip is fulfilling. Notice that (I1)-(I3) have been ordered resorting solely to what I've called the vertical dimension.

Now think of the hypothetical scenario presented by Jacob and Jeannerod (2005, p. 22), introduced in chapter 1 (section 3.2.4). They talk of the non-basic action of turning on the light, which they claim can be brought about by what they first term the basic action of pressing a switch, and by what they subsequently term the basic action of pressing a switch with one's right index finger. As it was with broadly congruent and strictly congruent mirror neurons, also in this case we can order these three action descriptions hierarchically as follows:

(A1') turning the light;
(A2') pressing the switch;
(A3') pressing the switch with one's right index finger.

A grasping with the hand using a precision grip, therefore, can be described at least at the above highlighted three different levels. These are ordered hierarchically in such a way that turning on the light at level (A1') is means-end generated by pressing the switch at level (A2'), and pressing the switch at level (A2') is means-end generated by pressing the switch with one's right index finger at level (A3'). Once more, I'm leaving it open that there could be one or more action descriptions that are means-end-generationally intermediate between, say, pressing the switch and pressing the switch with one's right index finger—such as pressing the switch with one's hand. Again, means-end generation applies insofar as (A1'), (A2') and

(A3') occupy the same time interval[55]. It is important to note that different kinds of means-end-generation are involved between (A1') and (A2') on the one hand and between (A2') and (A3') on the other hand. Pressing the switch (at level (A2')) is a means to turning on the light (at level (A3'))—an outcome that could likewise be achieved by opening the curtains, if there's light outside. By contrast, pressing the switch (at level (A2')) and pressing the switch with one's right index finger (at level (A3')) are related by that species of means-end-generation that involves determination.

The following intentions may be specified as underlying (A1'), (A2') and (A3'), respectively:

(I1') the intention to turn on the light;
(I2') the intention to press the switch;
(I3') the intention to press the switch with one's right index finger.

This time, only the intention to press the switch (at level (I2')) and that to press the switch with one's right index finger (at level (I3')) are motor intentions, given that their content involves a series of bodily configurations, and it imposes constraints on which bodily configurations are going to count as the fulfillment of the intention to press the switch. Suppose that one managed to turn on the light by performing some sort of grasping action on the switch: that wouldn't count as pressing the switch, insofar as the two actions, grasping and pressing, are characterized in terms of different bodily configurations. By contrast, the intention

[55] At least, under an understanding of turning on the light that is means-end-generated, rather than caused by, pressing the switch.

to turn on the light is one whose fulfillment doesn't require any specific series of bodily configurations: anything that is going to bring that result about will do, and the attainment of this outcome doesn't impose any constraints on the bodily movements involved. If there were a device that enabled you to turn on the lights by blinking, you'd be able to turn on the lights by blinking, instead of pressing a switch, or instead of opening the curtains. All these bodily realizations of the action of turning on the lights are too diverse to be grouped under an outcome that's defined solely in motor terms. To sum up, the action of pressing the switch with one's right index finger involves at least three intentions (the intention to turn on the light; the intention to press the switch; the intention to press the switch with one's right index finger) and, since no assumption is being made about the existence of intermediate levels of intention with respect to (I1')-(I3'), there could in principle be even more—such as the intention to press the switch with one's hand. Once again, (I1')-(I3') have been ordered by resorting solely to what I've called the vertical dimension.

The last result that I'm going to review exclusively in the light of the vertical dimension is the one by Gazzola and colleagues (2007), introduced in chapter 1 (section 3.2.5). In that study, two aplasic individuals born without hands or arms underwent an fMRI scan while they watched hand actions, and their brain activity was studied in comparison with that of sixteen typically developed individuals. Both aplasics, while observing hand actions, activated those areas in their brain—part of the mirror neuron system—that had previously been shown to be involved in the execution of foot or mouth actions. Aplasics are exposed to an action—grasping—that can be executed by different means—hands, mouth or feet. The following has to be borne in mind.

> While viewing hand actions, both TDs [i.e., typically developed individuals] and aplasics activated a combination of effector-unspecific areas and regions devoted to the effector that the observer would use to perform the observed action: the hand for TDs and the foot or mouth for aplasics.
>
> (Gazzola et al., 2007, pp. 1236-1237)

The action that aplasics observe can therefore be analysed as follows:

(A1'') grasping;
(A2'') grasping with the hand.

This action hierarchy can be analysed in an entirely analogous way to the earlier introduced one ((A1)-(A3)) involving grasping with different effectors, so I'm not going to dwell on the details of its analysis. Rather, I'd like to point out the difference between the intentions underlying (A1'')-(A2'') according to a TD agent on the one hand, and according to an aplasic observer on the other hand. In the first case, they are:

(I1''$_{TD}$) the intention to grasp

(I2''$_{TD}$) the intention to grasp with the hand.

In the second case, they are:

($I1''_{A}$) the intention to grasp;

($I2''_{A}$) the intention to grasp with the mouth or foot

The intentions listed at levels ($I1''_{TD}$)-($I2''_{TD}$) as those listed at levels ($I1''_{A}$)-($I2''_{A}$) are motor intentions, and it can be said that the TD agent has at least two intentions (($I1''_{TD}$)-($I2''_{TD}$)), and is recognized to have at least two intentions (($I1''_{A}$)-($I2''_{A}$)) by the aplasic observer. Note, however, that from the point of view of the aplasic observer the intention to grasp with the hand has been remapped onto the corresponding one in the aplasic's motor repertoire, namely the intention to grasp with the hand or foot. The interpretation of the study by Ferrari and colleagues (2005) is entirely analogous to that of Gazzola et al. (2007), so I'm going to leave it aside.

5.2 Re-interpreting results by resorting to the horizontal dimension only

Now recall the study by Fogassi and colleagues (2005), a single-cell recording study on the inferior parietal lobule (IPL) of the macaque monkey brain in which it is shown that grasping-for-eating is encoded differently from grasping-for-placing by mirror neurons. The authors suggest an interpretation of their findings according to which some of the recorded mirror neurons "not only code the observed motor act but also allow the observer to understand the agent's intentions" (Fogassi et al., 2005, p. 662), from which it would seem that they

conceive of eating and placing as the intentions of the acting individual, and, by contrast, it is not clear that the act of grasping has an intention associated to it.

The first thing to note is that, unlike the actions considered in the previous paragraph, the two actions involved in the experiment by Fogassi and colleagues can be segmented into temporally distinct components such as *grasp* on the one hand and *eat* on the other hand, or *grasp* on the one hand and *place* on the other hand. Due to this temporal segmentation, in order to analyze this action we're going to resort to what I've been calling a horizontal dimension for action description. The first thing to point out is that, out of the recorded neurons that coded for grasping, some responded differentially depending on the subsequent act, whereas others discharged in the same way regardless of the following segment of action (Fogassi et al., 2005, p. 662). In the light of this, it is possible to say that, while acting in the grasping in order to place condition, the agent has:

(i) the intention to grasp, and

(ii) (subsequently) the intention to place.

This is not the whole story, however, insofar as, on the basis of the fact that the firing of some grasping neurons is indeed modulated by the subsequent segment of action (eating or placing) (Fogassi et al., 2005, p. 662), it is possible to maintain that, in addition, to (i) and (ii), the agent also has:

(iii) the intention to grasp-in-order-to-place.

Note that (i), (ii) and (iii) all lie on the same level of a hierarchy.

At this point, you might want to challenge the addition of (iii) as redundant. However, I'm going to show that, far from being redundant, an interpretation featuring intentions that take into account the subsequent segment of the action rather provides the best explanation for the following result. An experiment on

typically developing vs. autistic children (Cattaneo et al., 2007) shows that a similar sequence of action segments may be executed and perceived as seamless or as detached. In this EMG recording, the activation of the mylohyoid muscle, which is employed in mouth opening and therefore while eating, was recorded both on TD children and on autistic children. The recording was carried out both while subjects executed and while they observed two kinds of actions—the same actions involved in the experiment by Fogassi and colleagues (2005): grasping to bring to the mouth and grasping to place. For TD children, it was found that:

(a) in execution phase, there was an increase of the EMG activity of the mylohyoid muscle as soon as an act of grasping in order to bring to the mouth began. By contrast, this activation was absent when the subject grasped an object in order to place it into a container.

(b) In observation phase, an immediate activation of the subject's mylohyoid muscle took place as soon as the experimenter grasped food to bring it to the mouth. No such activation was observed in relation to grasping in order to place into a container.

The authors interpret these findings as showing that the action of grasping in order to bring to the mouth is treated by the observer as a chain, a unified whole, that makes it possible for the mylohyoid muscle, which controls the last segment of the action, to be activated from the outset of the either executed or observed action (Cattaneo et al., 2007, p. 17829).

What about children with autism? Cattaneo and colleagues write:

> A behavior radically different from that of TD children was found in children with autism both during the execution and the observation of actions done by others. The most striking result was that during the

> execution of grasping to eat there was no activation of the MH muscle during the reaching and grasping phase. Its activation was found only during bringing to the mouth.
>
> (Cattaneo et al., 2007, p. 17829)

In a nutshell, while typically developing children seem to recognise the motor intention that lies further along in the sequence, autistic children are unable to do so, and only recognise the motor intention to bring an object to one's mouth towards the end of the sequence. This may be taken as showing that, while normally action segments can be performed and recognised as chains, sometimes recognition of this chain-like structure may break down, so that an individual can only perform (or understand) an action segment at a time. For our purposes, this may be taken to mean that, while grasping to bring to the mouth, the typically developing child is going to have all the three following intentions:

(i) the intention to grasp, and

(ii) (subsequently) the intention to bring to the mouth.

(iii) the intention to grasp-in-order-to-bring-to-the-mouth.

By contrast, the autistic child is likely to have only the former two.

5.3 Re-interpreting results by combining the vertical dimension with the horizontal dimension

Up to this point, we've discussed motor intentions related to different segments of an action, by considering segments displayed along just one level of the vertical

hierarchy. Now I'm going to discuss a way to use both dimensions in the interpretation of an experiment. There is an fMRI study (Iacoboni et al., 2005) which is often taken as parallel to Fogassi and colleagues' (2005) result. In this study, subjects were exposed to the following three kinds of stimuli:

a. *Actions:* clips showing grasping hand actions (grasping a mug either with a precision grip or with a whole-hand prehension) without a context;
b. *Contexts:* clips showing context alone (scenes in which objects were displayed, suggesting a *before tea* or *after tea* kind of situation);
c. *Intentions:* clips showing grasping hand actions (grasping a mug either with a precision grip or with a whole-hand prehension), embedded in a specific context, *before tea* or *after tea*, which, respectively, suggested the intentions *drinking* or *cleaning*.

The third kind of conditions—*intentions*—produced a significantly higher increase of the signal, compared with *actions* and *contexts*, in areas—the posterior part of the inferior frontal gyrus and the adjacent sector of the ventral premotor cortex—thought to contain mirror neurons (Iacoboni et al., 2005, p. 529). Iacoboni and colleagues' interpretation is that:

> premotor mirror neuron areas—areas active during the execution and the observation of an action—previously thought to be involved only in action recognition are actually also involved in understanding the intentions of others. To ascribe an intention is to infer a forthcoming new goal […].
> (Iacoboni et al., 2005, p. 529)

These statements are based on the understanding that mirror neuron activity in Iacoboni and colleagues' study is sensitive to the intentions of drinking vs. cleaning up rather than to the actions of grasping in different ways. But, if one takes into account both dimensions of explanation that I've proposed, this scenario can be complicated in at least two ways.

First of all, Iacoboni and colleagues' interpretation is based on a contrast between *actions* and *intentions*, or between *what* is being done and *why* it is being done, which my notion of motor intention has shown to be inadequate. Grasping is underlain by an intention as much as drinking or cleaning—rather, there might be an issue in assigning a level in the vertical hierarchy to the respective intentions. In particular, I claim that *cleaning* belongs to levels higher up in the vertical dimension with respect to *grasping*, insofar as the latter, but not the former, involve a kind of motor content—no constraint on the relevant sequence of bodily movements holds for cleaning. If there really was to be a parallel between this result and Fogassi and colleagues' one, this would mean that mirror neuron activity is sensitive to something *moving the cup somewhere* (say, to the sink), rather than to cleaning. I don't think that Iacoboni's result by itself enables us to disentangle between these two available characterizations, and its moral should be determined on independent grounds—e.g., on the grounds that motor outcomes, and not intentions underlying actions higher up in the vertical dimension, are the sort of things that mirror neuron activity correlates with.

Secondly, there is a potential issue with the time interval corresponding to the fulfilment of the intentions of drinking and cleaning. The action of cleaning is subject to two potential interpretations. In the previous paragraph, I've been assuming that the action of cleaning *follows* the action of grasping. But an

alternative interpretation might have it that the action of cleaning *encompasses* that of grasping: in addition to being located higher up in the hierarchy, it can also be seen as more extended along the horizontal dimension. Call the action of cleaning that's less extended in time *cleaning*$_1$ and that which is more extended in time *cleaning*$_2$. Since, according to Iacoboni and colleagues (2005, p. 529), "[to] ascribe an intention is to infer a forthcoming new goal", they interpret the intentions of cleaning as being fulfilled *after* the cup has been grasped—and therefore they interpret *cleaning* as *cleaning*$_1$. Still, the results as they stand do not enable us to adjudicate between these two interpretations. There is, then, a variety of intentions with which mirror neuron activity could in principle correlate in the cleaning scenario:

(i) The intention to grasp;

(ii) The intention to grasp in order to move away (say, to the sink);

(iii) The intention to grasp in order to clean$_1$;

(iv) The intention to clean$_2$.

All the results that I've reviewed can now be seen to be open to more complex interpretations than previously available.

Conclusion: motor intentions—not so high, not so low

The central question of this thesis was: how are actions connected with intentions? In order to answer this question, first I needed to make it clear what an intention is. In the course of chapter 2, I've established that intentions are mental states representing action outcomes with a world-to-mind direction of fit, that they enable effective planning, and are inserted in a specific web of norms that differentiates them from other propositional attitudes—specifically desires, which have the same direction of fit, and that they are endowed with a content that is accessible to introspection. In chapter 3, I've highlighted the existence of *motor representations*, namely of representations of an action outcome that depict the self in action as a generator of forces, and thus determine the pattern of movements that are going to be executed, thereby driving action execution. Given the proximity of motor representations to action execution, an answer to the question as to how actions are connected with intentions was bound to lie somewhere at the intersection between the notion of intention and that of motor representation.

Chapters 2 and 3 highlighted a commonality between intentions and motor representations, namely the representation of an outcome. Chapter 3 provided the foundations for the notion of *motor intention*, a kind of intention whose content has relevant commonalities with that of a motor representation. Chapter 4 made it more precise what this commonality in terms of content consists in. On the one hand, it emphasized that the content of a motor intention, as well as that of a motor intention, involves a sequence of bodily configuration. On the other hand, it

suggested the possibility to divide motor representations at least into two categories:

(i) a category of representations whose content is likely to consist in a motor image, and

(ii) a category of representations whose content is likely to be captured in propositional terms.

The motor intentions corresponding to (i) refer to action outcomes by deferring to the former kind of motor representation, whereas the motor intentions corresponding to (ii) are likely to have the same content as that of the corresponding motor intentions. Thus we can see how motor intentions connect intentions with actions by, in some cases, deferring to a motor representation, which then drives action execution, and in other cases by having the same content as that of the corresponding motor representation.

Chapter 5 showed the independence of the characterization of motor intentions, and of intentions more generally, of temporal considerations, thus showing the inadequacy of the notions of *prior intention* and of *intention in action*. Thus I've shown the shortcomings of a previous attempt to account for the connection between intentions and actions, namely the notion of *intention in action*. Chapter 6 completed the picture by showing how motor intentions and non-motor intentions can be ordered in a hierarchy that features both a vertical and a horizontal dimension. Motor intentions can be located at the lower levels along the vertical dimension.

The results about mirror neurons presented in chapter 1 gave us the opportunity to explore the notion of a *motor outcome* (chapter 1, section 5), an outcome whose content at the same time involves reference to bodily movements

while enjoying a relative degree of independence and abstraction with respect to fine movement details. In the light of the notion of motor intention, it becomes plausible to suggest that, since mirror neuron activation correlates with motor outcomes, and motor outcomes are represented by motor intentions, then mirror neuron activation correlates with motor intentions. On the basis of this idea, and taking into account also the framework for intentions worked out in chapter 6, I've concluded my thesis by re-interpreting a series of experimental results concerning mirror neurons in the light of the notion of motor intention.

Now, where do we go from here? Consider the problem, sketched at the outset, of understanding other's actions. Suppose I'm in the presence of other people who are carrying out some familiar activities—say, someone's bringing food to her mouth, while someone else is moving objects away from table. While I observe these individuals, I understand what they are doing—that is, I understand that the goals of their actions are, respectively, eating and tidying up. What does this understanding consist in? Is it the case that I explicitly reason on the details of the action that I observe to then conclude that it's an instance of, e.g., eating? Or does my action understanding consist in a process akin to perception?

Note that we now have the theoretical tools for framing the fact that that any token action may simultaneously fulfil many intentions. For instance, my action of picking up objects from a table and storing them away in the cupboard may fulfil my intention to move an object, that to store an object away, as well as, simultaneously, the intention to obey my personal love of neatness—and these are hierarchically ordered. An observer may well be able to tell that I'm picking up objects and storing them away, and also that I'm tidying up the kitchen, while not necessarily understanding my action as fulfilling a more abstract kind of intention

such as that of obeying my personal love of neatness (see Jacob & Jeannerod, 2005; Jacob, 2008). On the basis of this example, we could think that action understanding can take place at many levels, from more concrete ones—e.g., being able to tell an act of grasping from an act of sweeping—to more abstract ones—e.g., being able to distinguish a tidying up just for one's own sake as opposed to tidying up in order to obey one's love of neatness. Could it be the case that, depending on what level of action understanding we are considering, the processes in which action understanding consists may significantly vary? If so, could it be the case that whether mirror neuron activity plays a role in action understanding depends on the level of action understanding we are considering?

It can be suggested, for instance, that action understanding at the level of motor intentions should be characterized differently with respect to action understanding at the level of non-motor intentions. A development of this hypothesis may for instance lead to circumscribing the role of mirror neurons in action understanding to a specific kind of action understanding (see Jacob, 2008). Finally, disentangling these issues would help clarifying how mindreading—namely the set of capacities involved in understanding others and predicting their behaviour (Stich & Nichols, 2003)—works, by shedding light on the variable levels of complexity that it might involve depending on the task at hand (see Goldman, 2006, 2009*b*, and de Vignemont, 2009 on the notions of high-level vs. low-level mindreading). In short, the present work, and what it provides the foundations for, may be thought as further steps in the intricate task of elucidating how we understand the actions and, more to the point, the intentions of others.

References

Adams, F., & Steadman, A. (2004). Intentional action in ordinary language: core concept or pragmatic understanding?. *Analysis*, *64*(282), 173-181.

Aglioti, S., Goodale, M. A., & DeSouza, J. F. X., (1995). Size Contrast Illusions Deceive the Eye but not the Hand. *Current Biology, 5,* 679-85.

Anscombe, G. E. M. (1957/1963). *Intention.* Harvard University Press. Second edition, Harvard University Press. Page references to the latter.

Avenanti, A., Bueti, D., Galati, G., & Aglioti, S. M. (2005). Transcranial magnetic stimulation highlights the sensorimotor side of empathy for pain. *Nature Neuroscience*, *8*, 955–960.

Bach, K. (1978). A Representational Theory of Action. *Philosophical Studies*, 34, 361-379.

Baron-Cohen, S. (1995). *Mindblindness: an essay on autism and theory of mind.* MIT Press/Bradford Books.

Becchio, C., Sartori, L., Bulgheroni, M., & Castiello, U. (2008). The case of Dr. Jekyll and Mr. Hyde: A kinematic study on social intention. *Consciousness and Cognition, 17* (3), 557-564.

Bekkering, H., Wohlschlager, A., & Gattis, M. (2000). Imitation of gestures in children is goal directed. *Quarterly Journal of Experimental Psychology A, 53,* 153–164.

Bizzi, E., Kalil, R. E., & Tagliasco, V. (1971). Eye–head coordination in monkeys. Evidence for centrally patterned organization. *Science, 173,* 452–4.

Brand, M. (1984). *Intending and Acting*. Cambridge, MA: MIT Press.

Bratman, M. (1984). Two faces of intention. *The Philosophical Review*, *93*(3), 375-405.

---------- (1987). *Intention, Plans, and Practical Reason.* Cambridge, MA: Harvard University Press.

Buccino, G., Binkofski, F., Fink, G.R., Fadiga, L., Fogassi, L., Gallese, V., Seitz, R.J., Zilles, K., Rizzolatti, G., & Freund, H.-J. (2001). Action observation activates premotor and parietal areas in a somatotopic manner: an fMRI study. *European Journal of Neuroscience, 13,* 400-404.

Buccino, G., Lui, F., Canessa, N., Patteri, I., Lagravinese, G., Benuzzi, F., Porro, C.A., & Rizzolatti, G., (2004). Neural circuits involved in the recognition of actions performed by nonconspecifics: an fMRI study. *Journal of Cognitive Neuroscience, 16,* 114–126.

Buekens, F, Maesen, K., & Vanmechelen, X. (2001). Indexicaliteit en dynamische intenties. *Algemeen Nederlands Tijdschrift voor Wijsbegeerte, 93,* 165-180.

Butterfill, S. A., & Sinigaglia, C. (2012). Intention and Motor Representation in Purposive Action. *Philosophy and Phenomenological Research*.

Cattaneo, L., Fabbri-Destro, M., Boria, S., Pieraccini, C., Monti, A., Cossu, G., & Rizzolatti, G. (2007). Impairment of actions chains in autism and its possible role in intention understanding. *Proceedings of the National Academy of Sciences*, *104*(45), 17825-17830.

Clark, A. (2001). Visual Experience and Motor Action: Are the Bonds Too Tight? *Philosophical Review, 110*(4), 495-519.

Clark, M. A., Merians, A. S., Kothari, A., Poizner, H., Macauley, B., Rothi, L. J. G., et al. (1994). Spatial planning deficits in limb apraxia. *Brain, 117,* 1093–106.

Csibra, G. (2007). Action mirroring and action understanding: an alternative account. In Patrick Haggard, Yves Rossetti and Mitsuo Kawato (eds.), *Sensorimotor foundations of higher cognition.* Oxford University Press.

Csikszentmihaly, M. (1990). *Flow: The Psychology of Optimal Experience.* New York: Harper and Row.

---------- (1997). *Finding Flow: The Psychology of Engagement with Everyday Life.* New York: Basic Books.

Danto, A. C. (1965). What we can do. *American Philosophical Quarterly, 2*(2), 141-148.

Danto, A. C., & Morgenbesser, S. (1963). What We Can Do. *The Journal of Philosophy, 60*(15), 435-445.

Davidson, D. (1963). Actions, reasons and causes. *Journal of Philosophy*, 60, 685–700.

---------- (1973). Freedom to act. In T. Honderich (ed.), *Essays on Freedom of Action*. London: Routledge and Kegan Paul, 137–56. (Reprinted in Davidson, 1980, 63–81.)

---------- (1978). Intending. In *Essays on Actions and Events*, Oxford: Oxford University Press, 1980, pp. 83–102.

---------- (1980). *Essays on Actions and Events*. Oxford: Oxford University Press

De Jaegher, H. (2009). Social understanding through direct perception? Yes, by interacting. *Consciousness and Cognition*, *18*(2), 535-542.

Di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., & Rizzolatti, G. (1992). Understanding motor events: a neurophysiological study. *Experimental Brain Research, 91,* 176-80.

de Vignemont, F. (2009). Drawing the boundary between high-level and low-level

mindreading. Philosophical Studies, 144(3), 457-466.

Dretske, F. (1981). Knowledge and the flow of information. Cambridge, MA: MIT Press.

Evans, G. (1982). *The Varieties of Reference.* Oxford: Oxford University Press.

Ferrari, P. F., Rozzi, S., & Fogassi, L. (2005). Mirror neurons responding to observation of actions made with tools in monkey ventral premotor cortex. *Journal of Cognitive Neuroscience*, *17*(2), 212-226.

Fourneret, P., & Jeannerod, M. (1998). Limited conscious monitoring of motor performance in normal subjects. *Neuropsychologi*a, 36, 1133-1140.

Frankfurt, H. G. (1978). The problem of action. *American Philosophical Quarterly*, 157-162.

Gallagher, S. (2001). The practice of mind. Theory, simulation or primary interaction?. *Journal of Consciousness Studies*, *8*(5-7), 5-7.

---------- (2004). Understanding interpersonal problems in autism: Interaction theory as an alternative to theory of mind. *Philosophy, Psychiatry, & Psychology*, *11*(3), 199-217.

Gallese, V. (2001). The shared manifold hypothesis. From mirror neurons to empathy. *Journal of consciousness studies*, *8*(5-7), 5-7.

---------- (2003) A neuroscientific grasp of concepts: From control to representation. *Philosophical Transactions of the Royal Society. London B., 358,* 1231-1240.

Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain, 119,* 593-609.

Gallese, V., & Goldman, A.I. (1998). Mirror neurons and the simulation theory of mindreading. *Trends in Cognitive Sciences, 2,* 493–501.

Gallese, V., & Lakoff, G. (2005). The brain's concepts: The role of the sensory-motor system in conceptual knowledge. *Cognitive neuropsychology*, *22*(3-4), 455-479.

Gallese, V., Keysers, C., & Rizzolatti, G. (2004). A unifying view of the basis of social cognition. *Trends in Cognitive Sciences, 8*, 396-403 .

Gazzola, V., van der Worp, H., Mulder, T., Wicker, B., Rizzolatti, G., & Keysers, C. (2007). Aplasics Born without Hands Mirror the Goal of Hand Actions with Their Feet. *Current Biology, 17,* 1235–1240.

Gergely, G., Bekkering, H., & Király, I. (2001). Rational imitation of goal-directed actions in 14-month-olds. In *Proceedings of the Twenty-third Annual Conference of the Cognitive Science Society* (pp. 325-330). Lawrence Erlbaum.

Goldman, A. I. (1970). *A Theory of Human Action*. Princeton University Press.

---------- (1989). Interpretation psychologized. *Mind and Language, 4*(3), 171–185.

---------- (1992). Empathy, mind and morals. *Proceeding and Addresses of the American Philosophical Association*, *66*(3), 17–41. Reprinted in M. Davies, & T. Stone (Eds.), 1995. *Mental Simulation* (pp. 185– 208). Oxford: Blackwell.

Goldman, A. (2000). The mentalizing folk, in *Metarepresentation* (Sperber, D., ed.), Oxford University Press.

---------- (2006). *Simulating Minds: The Philosophy, Psychology, and Neuroscience of Mindreading*. New York: Oxford University Press.

---------- (2008). Mirroring, mindreading, and simulation. In J. A. Pineda (Ed.), *Mirror neuron systems: The role of mirroring processes in social cognition* (pp. 311–330). New York: Humana Press.

---------- (2009*a*). Mirroring, simulating and mindreading. *Mind and Language, 24*(2), 235– 252.

---------- (2009*b*). Replies to Perner and Brandl, Saxe, Vignemont, and Carruthers. *Philosophical studies*, *144*(3), 477-491.

Gopnik, A., & Meltzoff, A. (1997) *Words, Thoughts, and Theories*. Cambridge, Mass.: Bradford Books/MIT Press.

Gopnik, A., & Wellman, H. (1992). Why the child's theory of mind really *is* a theory. *Mind and Language*, *7*, 145–171.

Gordon, R. (1986). Folk psychology as simulation. *Mind and Language, 1*(2), 158–171.

---------- (1995). Sympathy, simulation, and the impartial spectator. *Ethics*, *105*(4), 727– 742.

Grafton, S. T., Arbib, M. A., Fadiga, L., & Rizzolatti, G. (1996). Localization of grasp representations in humans by positron emission tomography. *Experimental Brain Research*, *112*(1), 103-111.

Grafton, S. T., & Hamilton, A. (2007). Evidence for a distributed hierarchy of action representation in the brain. *Human Movement Science, 26*(4), 590–616.

Graziano, M. S. (2010). Ethologically Relevant Movements Mapped on the Motor Cortex. *Primate neuroethology*, 454.

Graziano, M. S., & Aflalo, T. N. (2007). Mapping behavioral repertoire onto the cortex. *Neuron*, *56*(2), 239-251.

Grèzes, J., & Decety, J. (2002). Does visual perception afford action? Evidence from a neuroimaging study. *Neuropsychologia, 40,* 1597-1607.

Haffenden, A. M., & Goodale, M. A. (1998). The effect of pictorial illusion on perception and prehension. *Journal of Cognitive Neuroscience, 10,* 122–136.

Hamilton, A., & Grafton, S. T. (2007). The motor hierarchy: from kinematics to goals and intentions. In P. Haggard, Y. Rossetti, & M. Kawato (Eds.), *Attention & Performance 22. Sensorimotor Foundations of Higher Cognition Attention and Performance* (pp. 381–408). Oxford: Oxford University Press.

Harman, G. (1976). Practical reasoning. *The Review of Metaphysics*, 431-463.

Hawthorne, J. (2004). *Knowledge and lotteries*. Oxford: Clarendon.

Heyes, C. (2010). Where do mirror neurons come from? *Neuroscience & Biobehavioral Reviews*, *34*(4), 575-583.

Iacoboni, M., Molnar-Szakacs, I., Gallese, V., Buccino, G., Mazziotta, J.C., & Rizzolatti, G. (2005). Grasping the intentions of others with one's own mirror neuron system. *PLoS Biology, 3,* 529-535.

Iacoboni, M., Woods, R. P., Brass, M., Bekkering, H., Mazziotta, J. C., & Rizzolatti, G. (1999). Cortical mechanisms of human imitation. *Science*, *286*(5449), 2526-2528.

Israel, D., Perry, J., & Tutiya, S. (1993). Executions, motivations, and accomplishments. *The Philosophical Review*, 515-540.

Jacob, P. (2008). What do mirror neurons contribute to human social cognition? *Mind & Language, 23,* 190–223.

---------- (2009). A Philosopher's Reflections on the Discovery of Mirror Neurons. *Topics in Cognitive Science, 1*(3), 570-595.

Jacob, P., & Jeannerod, M. (2005). The motor theory of social cognition: a critique. *Trends in Cognitive Sciences, 9,* 21-25.

Jeannerod, M. (1981). Intersegmental coordination during reaching at natural visual objects. *Attention and performance IX (ed. J. Long and A. Baddeey)*, *9*, 153-168. Erlbaum, Hillsdale, NJ.

---------- (1986). The formation of finger grip during prehension, a cortically mediated visuomotor pattern. *Behavioural Brain Research, 19,* 99-116.

---------- (1994). The representing brain: Neural correlates of motor intention and imagery. *Behavioral and Brain Sciences, 17*(2), 187–245.

---------- (2006). *Motor Cognition: What Actions Tell the Self.* Oxford University Press.

Kahneman, D. and Tversky, A. (1982) The simulation heuristic, in *Judgment Under Uncertainty* (Kahneman, D., Slovic, P. and Tversky, A., eds), pp. 201–208, Cambridge University Press.

Keysers, C., Wicker, B., Gazzola, V., Anton, J–L., Fogassi, L., & Gallese, V. (2004). A touching sight: SII/PV activation during the observation of touch. *Neuron*, 42: 335–346.

Knobe, J. (2006). The Concept of Intentional Action: A Case Study in the Uses of Folk Psychology. *Philosophical Studies, 130,* 203-231.

Kosslyn, S. M., Pinker, S., Smith, G. E., & Schwartz, S. P. (1979). On the demystification of mental imagery. *Behavioral and Brain Sciences* 2, 535-82.

Kurby, C. A. & Zacks, J. M. (2008). Segmentation in the perception and memory of events. *Trends in Cognitive Sciences, 12*(2), 72-79.

Landauer, T. K. (1962). Rate of implicit speech. *Perceptual and Motor Skills, 15*, 646.

Lashley, K. S. (1917). The accuracy of movement in the absence of excitation from the moving organ. *American Journal of Physiology, 43,* 169–94.

---------- (1951). The problem of serial order in behavior. In *Cerebral mechanisms and behavior* (ed. L. A. Jeffress), pp. 112–36.Wiley, New York.

Leslie, A. M. (1994*a*). Pretending and believing: Issues in the theory of ToMM. *Cognition*, *50*, 211–238 .

Lhermitte, F. (1983). Utilisation behaviour and its relation to lesions of the frontal lobes. *Brain, 106,* 237–255.

Liepmann, H. (1905). *Ueber Störungen des Handelns bei gehirnkrankin*. S. Kargen, Berlin.

Malle, B. F. (2006). The moral dimension of people's intentionality judgments. *Journal of Culture and Cognition*, *6*, 87-112.

Malle, B. F., & Nelson, S. E. (2003). Judging mens rea: The tension between folk concepts and legal concepts of intentionality. *Behavioral sciences & the law*, *21*(5), 563-580.

McDowell, J. (1994/1996). *Mind and World: With a New Introduction.* Harvard University Press.

Mele, A. (2001). Acting intentionally: Probing folk notions. In B. F. Malle, L. J. Moses, & D. A. Baldwin (Eds.), Intentions and intentionality (pp. 27–43). Cambridge, MA: MIT Press.

Mele, A., & Moser, P. (1994). Intentional action. *Noûs, 28*, 39-78.

Meltzoff, A. N. (1988). Infant imitation and memory: Nine-month-olds in immediate and deferred tests. *Child development*, 217-225.

Milner, A. D., & Goodale, M. A. (1995). *The Visual Brain in Action*. Oxford: Oxford University Press.

Mott, F.W., & Sherrington, C. S. (1885). Experiments upon the influence of sensory

nerves upon movement and nutrition of the limbs. *Proceedings of the Royal Society B: Biological Sciences, 57,* 481–8.

Mukamel, R., Ekstrom, A. D., Kaplan, J., Iacoboni, M., & Fried, I. (2010). Single-Neuron Responses in Humans during Execution and Observation of Actions. *Current Biology, 20*(8), 750-756.

Murata, A., Fadiga, L., Fogassi, L., Gallese, V., Raos, V., & Rizzolatti, G. (1997). Object representation in the ventral premotor cortex (area F5) of the monkey. *Journal of neurophysiology*, *78*(4), 2226-2230.

Nadelhoffer, T. (2005). Skill, luck, control, and intentional action. *Philosophical Psychology*, *18*(3), 341-352.

Newtson, D. (1973). Attribution and the unit of perception of ongoing behavior. *Journal of Personality and Social Psychology, 28*(1). 28-38.

Ochipa, C., Rapcsack, S. Z., Maher, L.M., Rothi, L. J. G., Bowers, D. and Heilman, K. M. (1997). Selective deficit of praxic imagery in ideomotor apraxia. *Neurology, 49,* 474–80.

O'Shaugnessy, B. (1991). *Searle's theory of action*. In E. Lepore and R. Van Gulick (eds.), *John Searle and his Critics*. Cambridge, MA: Blackwell, 271-287.

Pacherie, E. (2000). The content of intentions. *Mind and Language*, *15*(4), 400–432.

---------- (2002). Intention. *Encyclopedia of Cognitive Science*. London: MacMillan, 599-604.

---------- (2003). La dynamique des intentions*. Dialogue, XLII, 3*, 447-480.

---------- (2006). Towards a dynamic theory of intentions. In S. Pockett, W.P. Banks & S. Gallagher (eds) *Does Consciousness Cause Behavior? An Investigation of the Nature of Volition.* Cambridge, MA: MIT Press, 145-167.

---------- (2008). The phenomenology of action: A conceptual framework. *Cognition, 107*(1), 179-217.

---------- (2011). Nonconceptual representations for action and the limits of intentional control. *Social Psychology*, *42*(1), 67.

---------- (ms.). Action concepts.

Pacherie, E., & Haggard, P. (2010). What are intentions? *Benjamin Libet and Agency*, 70-84.

Paivio, A. (1986). *Mental representations: A dual coding approach.* Clarendon Press.

Pitt, D. (2012). Mental Representation. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy.*

Pylyshyn, Z. (1984). *Computation and Cognition.* MIT Press.

Ramenzoni, V. C., Davis, T. J., Riley, M. A., Shockley, K., & Baker, A. A. (2011). Joint action in a cooperative precision task: nested processes of intrapersonal and interpersonal coordination. *Experimental Brain Research 211*, 447–457.

Rizzolatti, G., Camarda, R., Fogassi, L., Gentilucci, M., Luppino, G., & Matelli, M. (1988). Functional organization of inferior area 6 in the macaque monkey: II. Area F5 and the control of distal movements. *Experimental Brain Research*, *71*(3), 491-507.

Rizzolatti, G., Fadiga, L., Gallese, V., & Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Cognitive Brain Research, 3,* 131-141.

Rizzolatti, G., Fogassi, L., & Gallese, V. (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. *Nature Reviews Neuroscience, 2,* 661-670.

---------- (2002). Motor and cognitive functions of the ventral premotor cortex. *Current opinion in neurobiology*, *12*(2), 149-154.

Rizzolatti, G., & Sinigaglia, C. (2008). *Mirrors in the Brain: How Our Minds Share Actions and Emotions.* Oxford University Press: Oxford.

---------- (2010). The Functional Role of the Parieto-frontal Mirror Circuit: Interpretations and Misinterpretations. *Nature Reviews Neuroscience, 11,* 264-274.

Rossetti, Y., Pisella, L., & Vighetto, A. (2003). Optic ataxia revisited: visually guided action versus immediate motor control. *Experimental Brain Research 153*, 171–179.

Searle, J. (1983). *Intentionality: An Essay in the Philosophy of Mind.* New York, Cambridge University Press.

---------- (1991). *Response: The Background of Intentionality and Action.* In E. Lepore and R. Van Gulick (eds.), *John Searle and his Critics*. Cambridge, MA: Blackwell, 289-299.

Setiya, K. (2010). Intention. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy.*

Shallice, T. (1988). *From neuropsychology to mental structure.* Cambridge: Cambridge University Press.

Singer, T., Seymour, B., O'Doherty, J., Kaube, H., Dolan, R., & Frith, C. (2004). Empathy for pain involves the affective but not sensory components of pain. *Science*, *303*, 1157–1162.

Slachewsky, A., Pillon, B., Fourneret, P. Pradat-Diehl, Jeannerod, M., & Dubois, B. (2001). Preserved adjustment but impaired awareness in a sensory-motor conflict following prefrontal lesions. *Journal of Cognitive Neuroscience*, 13, 332-340.

Sperber, D. (2004). *'Mirror neurons' or 'concept neurons'?* Available at: http://www.interdisciplines.org/medias/confs/archives/archive_8.pdf. Accessed on February 10th 2013.

Stich, S., & Nichols, S. (2003). "Folk Psycholgy," in *The Blackwell Guide to Philosophy of Mind,* S. Stich and T. Warfield (eds.), Oxford: Blackwell.

Thomas, N.J.T. (2010). Mental Imagery. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy.*

Tucker, M., & Ellis, R. (1998). On the relations between seen objects and components of potential actions. *Journal of Experimental Psychology: Human perception and performance*, *24*(3), 830.

Uithol, S., van Rooij, I., Bekkering, H., & Haselager, P. (2011). Understanding motor resonance. *Social neuroscience*, *6*(4), 388-397.

---------- (2011). What do mirror neurons mirror?. *Philosophical Psychology*, *24*(5), 607-623.

---------- (2012). Hierarchies in action and motor control. *Journal of cognitive neuroscience*, *24*(5), 1077-1086.

Umiltà, M.A., Escola, L., Intskirveli, I., Grammont, F., Rochat , M., Caruana, F., Jezzini, A., Gallese, V., & Rizzolatti, G. (2008). How pliers become fingers in the monkey motor system. *Proceedings of The National Academy of Sciences USA 105*(6), 2209–2213.

Velleman, D. (2007). What good is a will? In Anton Leist & Holger Baumann (eds.), *Action in Context.* de Gruyter/Mouton.

Wicker, B., Keysers, C., Plailly, J., Royet, J.–P., Gallese, V., & Rizzolatti, G. (2003). Both of us disgusted in my insula: The common neural basis of seeing and feeling disgust. *Neuron*, *40*, 655–664.

Wolpert, D. M., Doya, K., & Kawato, M. (2003). A unifying computational framework for motor control and social interaction. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *358*(1431), 593-602.

Wolpert, D., & Ghahramani, Z. (2000). Computational principles of movement neuroscience. *Nature Neuroscience, 3,* 1212–7.