

How voluntary are minimal actions ?

in S. Maasen, W. Prinz, J. Roth, *Voluntary action*, Oxford, Oxford University Press, 2003, 202-219.

Joëlle Proust
CNRS, Institut Jean-Nicod

This paper aims at exploring how intentional a piece of behavior should be to count as an action, and how a minimal view on action, not requiring a richly intentional causation, may still qualify as voluntary.

Knowing what counts as an action is not only a matter of nominal definition. It has important consequences in education, law, as well as in ethics. One unresolved question in the domain of action consists in determining which features, if any, are distinctive of human action as contrasted with non-human animal action. The interest of studying what Kent Bach (1978) called "minimal actions" is that they seem not to presuppose any prior intention, not even any conscious decision. Defining a minimal action as a guided bodily movement (not requiring an intentional - belief-desire - causation) will help us discuss some issues that become easily blurred within the folk-psychological approach to full-fledged, richly intentional actions. In the first section, I will first develop the contrast between the two kinds of action and offer reasons for taking the category of minimal action as a functionally relevant one. In the second section, this "austere" definition of action will be used to test our intuitions about what counts as a voluntary action. In ordinary usage, three different ways of explaining what makes an action voluntary can be found. The first involves the sense of agency which normally, but not necessarily, accompanies self-generated bodily movement. The second invokes the fact that the subject wanted to act as she did. The third, richer meaning of a voluntary action requires that an agent has a positive evaluation of what she did, in the sense that her second-order desires are in agreement with the action taken. We will examine how these three senses may apply to minimal actions. This analysis will hopefully contribute separating out causal, experiential and metaphysical considerations on voluntary action.

I - Full-blooded versus Minimal Actions

Standard philosophical approaches of action define it in terms of some psychological state causing a relevant bodily movement. While there is now a wide convergence on this causal approach of action, there are disagreements on the kind of psychological state deemed relevant for action. Davidson (1980) holds that what is causal is a *reason to act*, ie a "proattitude" (desire) relative to some state of affairs, in combination with the relevant belief that such and such a movement would bring about the desired condition. Goldman (1970) maintains similarly that the psychological states causing behavior are wants and beliefs. Searle (1983), Bratman (1987) and Mele (1992) claim that *intentions* form an indispensable link between reasons and actions. Intentions have two essential features : they guide behavior in virtue of their *representational content*; they move an agent to act in virtue of their *executive* properties.

What is important to note is that, in these various causal theories, an agent is not directly a cause of her actions; her beliefs, motivational or intentional states are. Also, these theories embed action in a *practical reasoning schema* : an agent has to judge that he desires P and that he believes Q in order to conclude that the action A recommends itself; she also has to assess the relative strength of her various competing reasons for performing an action of type A rather than of type B. In this way, the causes of an action coincide with the reasons the agent has for acting.

Thus the reason to act, in these standard theories, is both a *justification* of an action, and a *cause* of the corresponding physical behavior. In the intentional account, for example, a subject acts because she formed and was able to self-ascribe the corresponding intention. Forming an intention of a certain type and self-ascribing the corresponding intention further require from the subject an ability to use the relevant *concepts* in formulating her intentions, in appreciating her preferences and in

performing her practical reasoning. Many authors would also defend the view that an agent must have a concept of *self* to recognize the corresponding intentions as her own and take responsibility for the developing action. In this account of action, mental events and properties endowed with *conceptual* content play an overwhelming role in the very causation of an action.

Starting with this received view on action, there are two ways of relaxing the conditions for a particular behavior to qualify as an action. The first consists in disconnecting the causation of action from the appreciation of the reasons that justify the actions ; the second consists in questioning the importance of having a specific intention for an action to be triggered. Let us examine first why the definition of an action needs to be extended in these two respects.

A - Ability in justifying action as a precondition for acting

While it makes sense, biologically, that an organism should use all the knowledge available to it to control adequately its behavior, it is not clear that the elements of conceptual knowledge offered in rationalizing behavior should also be a *necessary* precondition for triggering an intention and the corresponding action. In other words, it seems plausible to allow a disconnection between performing an action of type A in context C, on the one hand, and rationalizing this performance, on the other hand.

This disconnection may be either radical or graded. In the radical approach, a non-linguistic animal would be attributed agency capacities just in case it performs some goal-directed behavior based on its prior encounters with similar situations. Such an organism may be said to act if it behaves on the basis of prior *learning* of what is to be done in context C (C includes endogeneous motivational states and exogeneous affordances), whether or not it engages in practical reasoning. Learning triggers the behavior in the sense that an organism exposed to motivating cues in a given context is disposed to perform an action of type A. Nevertheless learning is nothing else than a set of representations being activated in the control of behavior ; by using the internal indicators whose function is to signal meaningful events and properties (presence of a predator, of food, of a mate), the agent - human or non-human - can develop the adequate motor response. In this radical version of the cause/reason disconnection, an action may well develop without the agent consciously representing the situation in a space of reasons. Indeed learning can develop outside any conscious access to reasons for acting in this particular way.

The graded approach would simply admit the possibility of a loose connection between acting and rationalizing action. In this view, learning is a case in which reasons for acting might still be offered, although not necessarily by the agent himself. An observer would be allowed to substitute to the agent to explain in which way the chosen course of action conforms to the agent's best interests, given its other epistemic and motivational states. A reason to favour the graded approach has to do with the fact that reinforcement learning, in a way, *instantiates* a piece of practical reasoning. Reinforcement learning could be expressed through the following sequence of representations. Given that certain internal representations have been activated ("this is a predator"), given the established associative links between these representations and highly significant properties ("danger for self and for offspring"), the conclusion reached is that "flight (or protective attack) should be preferred". But the difference between this kind of embodied reasoning and full-blown practical reasoning is that reinforcement does not rely on a general and detached analysis of the facts at hand, as does practical reasoning. It does not need to explore counterfactual events and properties -- those which would not be present if the action was not to take place. Although reinforcement learning can be evaluated as a rational way of coping with a stable world, it cannot be equated with a form of practical reasoning because no representation of the alternative courses is needed in the former, while it is essential to the latter.

If the analysis above is on the right track, the very notion that a causal account of action should coincide with a justification in terms of reasons to act dissolves. A non-speaking animal or a young child cannot provide a justification for their actions, and cannot even represent to themselves the problem of justifying them. Not only because they cannot explain what they did to others or to themselves, cannot select the appropriate concepts while communicating the content of their intentions and therefore have no business with rationalizing what they did. But also because a non-speaking

animal or a young child cannot represent counterfactual situations, i.e. alternative courses of action available in the same context, with their consequences and their relative motivational strengths. This suggests that the radical view of the disconnection between action and reason-giving is more accurate than the graded view in the case of actions performed by non-speaking animals.

What holds for all kinds of animal or infant actions also holds for an extended set of human adult actions. There are many cases in which an agent has no interest in developing practical reasoning in its full-fledged form. The agent simply acts unreflexively, and not on the basis of a former representation, however sketchy, of the means-end relation at hand, and of her reasons to perform the corresponding action at the light of her best knowledge.

B - Intention as a precondition for acting

Another way of weakening the standard account of an action consists in questioning the essential role of intentions in initiating an action. Kent Bach boldly claimed that the supposition that every action should be taken as intentional - or at least, when unsuccessful, to involve an intentional part -, is "unsupported" (Bach, 1978, 363). By an *intention*, Bach understands a distinctive conscious piece of willing, a mental event that could be self-presented by the agent through the words "I will do A". Bach observes correctly that there are types of actions which are performed routinely and unthinkingly, actions for which it cannot be claimed that they were willed in the sense in which one willingly writes a letter or utters a sentence. Examples of these "minimal" actions are all kinds of postural or preattentive movements, such as scratching an itch, doodling, brushing a fly, avoiding an object (Bach, 1978), shifting a gear, pacing about the room, (Searle, 1983), impulsive or expressive actions such as tearing a photo into pieces out of rage, jumping up out of joy (Hursthouse, 1991).

Bach (1978) offers two main reasons for admitting non intentional kinds of actions. First, it seems difficult to reject from actions all the types of behavior cited above. They might well be the only kind of actions which infants, patients with executive disorders and "animals on the middle rungs of the phylogenetic ladder" are able to perform. Second, there is more to an action than its initiation. The way in which an action is carried out seems not to be determined at the level of a prior intention. Although the specific way in which an action is performed requires a form of awareness, the latter seems "below the level of intentions and reasons". The level of acting which is left when (conscious or unconscious) prior intentions are missing corresponds to what Bach calls a "minimal" action.

One might want to say that John Searle's analysis in his (1983) essay gives an adequate answer to Kent Bach's challenge. First, Searle acknowledges the existence of actions performed *without* prior intentions. In contrast with the situation where an agent forms a *priori* the intention to do A - an intention reportable by the linguistic form "I will do A" -, Searle claims that one often performs an action with no such prior intention, but rather with an occurrent thought of the linguistic form "I do A". When I scratch my back, I don't form the prior intention to the effect that I will scratch my back, I just do it (1983, 84). But there is no reason, according to Searle, to conclude that I do it so to speak unintentionally. This kind of mental state belongs to intentions, because it is constituted - and this is a central point - *by the conscious experience of acting*. According to Searle, it is "inseparable from action", and called for that very reason "an intention in action". (1983, 84)

Second, intentions in action have another feature which meets Bach's objection, that an intention should not only trigger an action, but control its development. When an agent has the intention to drive to her office, she does not form explicitly the prior intention to shift gears from second to third ; she does so as a result of her intention in action to shift gears from second to third. Such an intention in action has the function of specifying the details of how an intentional content - the goal of the corresponding action - should be realized. A further question here is to explain *how* this realization of a conceptual intentional content into a non-conceptual concrete behavior can be made possible. According to Searle, the explanation consists in embedding a second causal link in the intentional chain. The prior intention specifies the goal in conceptual terms ; the intention in action consists in an "experience of acting" which causes the specific bodily movements associated to it. In Searle's view, an experience of acting is a "presentation of its conditions of satisfaction" (Searle, 1983, 88). This ensures that the

representational character of the experience will be tightly linked to its distinctive "feeling" : being a *presentation* involves having a specific phenomenological character, an associated feeling of agency ; having *conditions of satisfaction* make it a fully intentional state, i.e. a state whose function is determined by a relationship between a mental representation and an external state of affairs.

Searle's notion of an intention in action certainly responds in part to Bach's problem. Indeed, it does account for the two essential constraints which make an action minimal. An intention in action has to do with the routine, unplanned ways of coping with the environment, whether physical or social ; it also accounts for the specific dynamics of the bodily movements through which an action is being performed.

Still there are features in Searle's analysis which conflict with the demands of minimal actions. First, the *direction of causation* of an intentional action is taken by Searle to be, in all cases, "mind to world"(Searle, 1983, 88) : an intention in action is a mental experience of trying which causes the bodily movement fitting that very intention, i.e. a bodily movement whose adequacy is an objective fact out in the world. In this analysis, an action whose direction of causation is world to world is a conceptual impossibility. Such a world to world direction of causation seems however involved each time an organism simply does what the context prompts him/her to do, i.e. each time a piece of goal-directed behavior is *stimulus-driven*. When for example a patient with an *imitation syndrom* is performing an action when she sees the same action performed by another agent, she is displaying an intentional behavior based, presumably, on the formation of intentions in action ; but the *causation* of her action is not mind to world for that. The world causes her to act, and she might be in a position to complain that people use her neurological problem to make her do things which she does not approve. A similar situation occurs in many cases in which normal agents are influenced to act as they do by contingent events in their environment : babies starting to cry when another baby cries, crowds of people running in panic to the same exit door, any one of us responding to a hand-shake approach by a hand-extension, etc. A last important set of less-than-intentional cases consists in what might be called deferred quasi-intentional behavior : an agent forms a plan, then forgets subsequently that he did, but nevertheless acts accordingly without a *conscious* intention in action when the context cues him for doing so. (See Wegner, Reader, vol. 2, 22 sq). To this category probably belongs the case of the post-hypnotic agent executing an action she does not remember having been instructed to perform.

What seems to make these cases difficult for Searle is 1) that his concept of an intention in action is tailored to be autonomously causative in action, that 2) that it *essentially* belongs to the family of conscious states and finally that 3) causation is effected through the very experience that the corresponding intention in action provides. What happens then when an action is performed *without* a conscious experience of doing so, for example in driving absent-mindedly ? Searle admits that,

"in such a case the intention in action exists without any experience of acting. The only difference then between them [= between this case and the case in which the action is conscious] is that the experience may have certain phenomenal properties that are not essential to the intention" (92).

In arguing in this way, Searle aims explicitly at finding a theoretical status for non conscious intention in action *symmetrical to* the case of non conscious perception in blindsight. A patient with blindsight has no phenomenal experience of seeing that P, but can still extract from vision at least some of its intentional properties. Similarly, a distracted person may have no phenomenal experience that he is driving, but rather have a non congruent experience (listening to his car radio, thinking about the elections, etc.) ; he may still accomplish the bodily movements which are part of the intentional conditions of satisfaction for driving.

This symmetry between perception and action is more apparent than real, however, and in particular it is easy to see that the parallel with blindsight does not work. For while a visual content can survive the absence of a visual experience, because the external object is what makes the extraction of a specific intentional content possible to begin with (Searle, 1983, 47), it does not seem open to us to say, at least within Searle's theory, that the experience of an action can go on causing the corresponding action when it has no phenomenological property. For in that case, there is no independent event - at least no event which Searle would be happy to consider - causing both the subjective experience and the bodily movement as is the case in perception (where one can be informationally connected with an object without

recognizing it). Given that an intention in action is essentially conscious, and is supposed to operate through the image of the action it provides, the notion of an unconscious intention in action remains rather obscure, and does not allow to understand how it might cause the action as if it had been conscious all along.

A more intelligible theory of the mental cause of an action would follow a functional rather than an experiential lead. There is indeed a mental process causing the action, on which the conscious experience of acting partly supervenes : it consists in activating a specific *motor representation*. Such a notion has all the features which make the concept of intention in action attractive, without excluding the possibility of minimal actions. Being both executive in its function and specific in its content, an active motor representation explains how an agent can execute in a concrete and flexible way a general plan ; it accounts for the adjustments that the agent has to make to reach the target state (by comparing dynamically the observed result with the anticipated goal). This solution further allows that a representation can be activated without being conscious or susceptible of becoming conscious. Final important merit : the activation of a motor representation does not presuppose that the source of the representational activation be endogeneous, i.e. constituted by some particular intention to act. In this general scheme, an action should not be defined in terms of its source, but in terms of its specific development from an internal model towards a goal with an appropriately monitored execution. What is pertinent is whether or not the bodily movements tend to be under the agent's guidance" : whatever the causal antecedents of a specific goal-directed movement may be, what makes it an action is the contribution of the corresponding agent to actively maintain the orientation of his bodily effort towards achieving a target event. In the case of minimal actions, the source of the action, i.e. the actual cause that triggers it, may be exogeneous (a motivating or "prepotent" stimulus) as well as endogeneous (a forgotten instance of planning, an overlearned routine). In this theory, minimal actions become the far end of a wide spectrum of behaviors ; the way in which an action is carried out may vary from a purely automatic, non-conscious process to a controlled and deliberate monitoring.

It is to the credit of Kent Bach to have sketched such a theory - which he calls Representational Causalism - well before neurophysiological findings could give this theory its empirical credentials (See Jeannerod, 1994, for a review). The central claim of Representational Causalism is that all actions - including minimal ones - "require "effective" representations for their initiation and execution" (Bach, 1978, 367). In this view, the representational character of the motor instructions allow the guidance to develop over time towards the goal. In other terms, acting presupposes essentially the ability to compare the representation of what is to be done with the various feedback representations conveyed by the agent's senses (what Bach calls "receptive representations"). Effective representations provide the motor commands, whereas the receptive representations correspond to the reafferences used as feedback for corrections or termination of the process. The set of effective and receptive representations, considered independently from their mind to world or world to mind "direction of fit", is aptly referred to as *executive representations*. Although Bach does not call his executive representations "motor representations", he insists that they are not propositional in form, and not necessarily linguistically coded ; nor do they necessarily imply any intentions or beliefs (1978, 366). Let us mention two other features that would deserve more attention, but can only be mentioned in the present article. Such a set of representations is stored in memory, and thus *allows* recall, mental simulation, as well as active role in executive feedforward and feedback. This choice of dispositional aspects of motor representations accounts for the role of imagery in simulation and in preparation of action, and for the fact that imagery, while useful, may also be entirely absent from performance. Furthermore, such representations may be activated in the absence of any conscious control or self-attribution, while again taking part in these processes when certain additional conditions are present.

Revising the constituents of a minimal action

As John Searle insists, "The key of Intentionality is conditions of satisfaction. It ought to be a rule in all of these discussions that nobody is allowed to talk about an

intentional phenomenon without telling us what its conditions of satisfaction are" (1991, 297). Now that we have shown that an action may develop without relying on prior pragmatic reasoning, or even on a conscious intention, it is time to make the conditions of satisfaction of a minimal action explicit.

As we saw, the causation may be world to world as well as mind to world ; therefore the *causal origin* of the activation of a specific motor representation should not be part of the definition of a minimal action. *A successful minimal physical action is such that a given motor representation takes control of behavior and monitors adequately occurrent feedback until its target event is reached.* As in the kind of analysis offered by Searle, this persistent role of the representation in the action, from triggering until completion, can be expressed as a kind of self-reference of the motor representation in the intentional content ; the latter can be articulated as :

that this very representation cause the appropriate bodily movement.

The "appropriateness" here is normally the result of prior learning, or alternatively, of hardwired responses to context, such as a flight induced by spotting a predator. Adequate monitoring is such that it makes gradually possible a transformation of the world in some expected condition through an appropriate sequence of active bodily movements. Inadequate monitoring would defeat the minimal action just as it defeats a full-fledged one. For example, if a minimal action consists in pacing about, tripping on the carpet or stepping on a brittle item would present a failure of the relevant motor representation in monitoring the input-output flow of information; new energy-consuming steps would have to be taken. If by scratching an itch, an extension of the itch or an open wound follows, these unfelicitous and unexpected consequences lead the agent to cope with a new situation. So minimal actions also have conditions of satisfaction ; failing to achieve the latter, or achieving them improperly, may interfere with other plans of the same agent.

Whereas a typical failure in full-fledged action consists in missing crucial steps or mistaking the target for another, it is less typical in the case of minimal actions, because they normally involve elementary bodily movements : launching one single automatic process, like walking or scratching, does not seem to lend itself to disruptive interferences, because being more automatic than the rest of the actions, it is likely to be successfully absorbed in the course of the other actions being launched in parallel. (One can scratch an itch, for example, while answering the phone). It may nevertheless sometimes happen that a minimal action never reaches its goal, maybe for lack of any adequate motor representation (e.g. scratching a part of the back where one cannot reach), for lack of time or for lack of other external conditions of satisfaction. Other ways in which a minimal action can fail to succeed is to fail to come to a stop by failing to register that its target event has been reached. This is the case of compulsive scratching or washing, for example, which can be analysed as cases of perseveration.

Let us take stock. What distinguishes executive representations from intentions as used in the standard causal theory is 1) that they are not necessarily conscious and 2) that they nevertheless actually guide the course of an action. The kind of relation they have to action, in other words, is functional rather than experiential. But all these welcome consequences are associated to a less than welcome one : there is now a gap to be bridged between conscious access to intention and action, on the one hand, and the actual representational causation of the action, on the other hand. There is moreover no guarantee that bridging this gap will leave our sense of agency unscathed. To understand why, we have to understand how representations become effective, i.e. causally efficacious in causing a bodily movement. We will ask this question in the context of a more general question, having to do with the ways in which it is justified - or only tempting, but wrong - to say that a minimal action is voluntary.

II - How voluntary are minimal actions ?

There are three senses at least in which an action is ordinarily called voluntary. The first has to do with the *feeling of will* associated with the action ; the second has to do with the *causal origin* of the action : an action is called voluntary in the latter sense if its cause is endogeneous and can be consciously accessed; the third sense involves *higher-order states* of the agent relative to the first-order intentional content of the action. Let us examine in turn each notion of will involved, in order to evaluate how

voluntary minimal actions can be.

A - Voluntariness as a feeling of willful action

Intuitively, speaking of a voluntary action suggests a specific feeling of being in charge, in contrast with cases when a move is triggered by physical forces or social constraints. In the case of minimal actions, an agent may also have a feeling of being in charge, even though she did not deliberate on his doing what she does. She has the sense of doing X voluntarily (jumping up from joy, scratching her arm, etc.). How is she aware of this specific quality of her behavior? A natural response would be to say that an agent gains this awareness on the basis of the specific motor representation presently activated. In this view, the motor representation would carry its own representational status on its sleeve, so to speak. If a representation adequately monitors current behavior, it would "tell" the subject *what* she does while she does it. Even in the specific case of minimal actions, where no prior intention is involved, an agent could recognize that she acts because whatever she does is guided by corresponding motor representations.

There are several reasons, however, to reject the general view that conscious awareness of agency is a byproduct of activating executive representations. First, neurological dissociations have been documented between awareness of having performed a movement and awareness of [thereby] performing a specific action. Clinical data from apraxia and from schizophrenia, as well as experimental results on normal subjects, suggest that a subject may know which action she is performing while being wrong about the way she is performing it, and vice versa: knowing which movement she executes, while being unaware of, or confabulating, the intentional motive of the movement. In other words, she may be able to characterize her goal and unable to report on the way she has acted towards it; reciprocally, she may be able to copy a particular gesture without being able to extract from it an information on the goal which this gesture normally subserves. In echopraxias, patients may also invent a goal for what seems to be a pure case of stimulus driven behavior. More surprisingly yet, an agent may identify an action through its goal and through the specific movement token which is used to reach it, without being able to identify correctly the author of this action.

What is largely documented in neuropsychological patients is also true for normal people in specific contexts. As every gym teacher knows, normal untrained children and adults are rather poor at reproducing accurately a specific token of a movement. Although they can describe exactly what they see being done, they seem to have trouble identifying the way in which they themselves are moving, in particular in the absence of a target object. In an experiment by Fournieret & Jeannerod (1998), normal subjects are asked to draw a sagittal line on a horizontal plane with a stylus, which they are not allowed to see directly. What they see is an image of their line-drawing on a monitor screen. Unknown to the subjects, the visual image is occasionally biased in deviating more or less from the line they actually produce; the subjects need to deviate their own movement in the opposite direction for achieving a sagittal line on the monitor. When asked to reproduce their drawing with eyes closed, right after each trial, the subjects tend to draw a line straight ahead, as suggested by the former visual reafference, in contradiction to any endogeneous source of information they might have had during their action.

These data on action awareness might be extended to the case of minimal actions, and suggest that in this case too, the perceptual *experience* of the action-related situation does not directly draw on the motor representation controlling the corresponding behavior. So it may be asked again: on the basis of what kind of information is a subject aware of her minimal actions as her own?

Normally, any mental or physical activity is perceived as originating in self or in some external event thanks to the information carried by a dedicated signal, telling whether a movement was effected by the individual; when absent, the brain would "interpret" a movement as unwilling, as when the body is passively subjected to some external force. The relevant signal is supposed to help compare reafferent signals with the signals that are expected on the basis of the current willed movement. It has been suggested that such a signal would be delivered by a mechanism underlying active perceptual activities, named corollary discharge, (Sperry, 1950) or efferent copy (von

Holst & Mittelstaedt, 1950). This mechanism is required to explain, for example, how ocular saccades can be taken into account and neutralized in interpreting visual input. When no such efferent copy signal is produced (one is sitting in a train), it is much more difficult to say whether oneself is moving or whether the perceived scene is. In our conscious sense of agency, a major component would thus consist in the *sense of effort* which is the subjective correlate of the corollary discharge of any action.

This sense of effort would account for the general feeling of having executed voluntarily an action, and would be lacking when a bodily movement happens involuntarily (by way of some external force, as when you're pushed by someone). An activation of this signal might certainly tell the agent, generally speaking, whether he is moving in a voluntary way. Yet this does not explain how the agent could become aware that he did such and such a *specific* type of movement or action.

The answer to this question might be related to the kind of expectations that are prompted to an agent when her behavior is controlled by a given motor representation. A subject might use the subpersonal information concerning the present input sequence as being actively gained and the feedback expectancies associated to a given motor representation (also activated at a subpersonal level). She would thus be able to perceive the intentional content of her action through the kind of visual feedback and end result which she is able (when her action is successful) to anticipate and experience sequentially in the visual, proprioceptive or auditory modes. In this view, a subject is able to perceive her own actions by looking directly at the dynamic visual, auditory, properties of the world. Perceiving an action would thus crucially depend on the very existence of a distinctive sequence of perceptual feedback, in particular of a visual nature. When this feedback is ambiguous or distorted, then the sense of what is done would be distorted accordingly.

Such a claim seems largely consonant with Daniel Wegner's view, according to which "we would tend to see ourselves as the authors of an act primarily when we had experienced relevant thoughts about the act at an appropriate interval in advance, and so could infer that our own mental processes had set the act in motion". The two approaches only seem to diverge on the inferential nature of the whole process. In the present view, while perceiving her own actions, an agent perceives a sequence of objects and properties in her environment and the transformations contingent to her bodily movements. This kind of perception is of a sensory kind ; it may give rise to a perceptual judgment as to the kind of action currently being performed, well before the action is completed. The action may be recognized perceptually through very tenuous cues without having to be inferred (no less than seeing a part of the Eiffel Tower leads one to *inferring* the presence of the Eiffel Tower ; the Eiffel Tower is perceived, and not inferred).

The fact that a subject can read off her own acting from her perceptual reafferences does not imply, however, that her perception is *reliable* in all circumstances : both approaches - Wegner's and the present one - converge again on this point. If the reafferences are either entirely unexpected by the agent - for lack of a prior experience or because the environment is very different from the ordinary, the subject may fail to perceive the action as her own (just as any pattern - visual, auditory, etc. - cannot be extracted if there is too much "noise").

If, on the other hand, as is the case with minimal actions, the reafferences are processed in an automatic way, with a diminished awareness, the sense of the action being voluntary should also be diminished. There might be numerous cases in which a subject admits that she executed an action "without realizing it", and pleads for distraction, or even denies having performed the action (as in many delusional states). Our definition nevertheless implies that if she exerted a control in performance, then she did perform the action.

One can object to this definition of a voluntary action through perceptual access, however, by claiming that it only emphasizes what is crucial for voluntary *movement*, while being silent about voluntary *action*. A subject can certainly *see* that she herself moved ; but finding out that she herself *acted* is more complicated, and calls for a second type of analysis.

B - Voluntariness as an endogeneous causal feature

At least in one sense of action, there is more to voluntary action than there is to

voluntary movement. An *action* being voluntary has to be contrasted with a *movement* being voluntary. I may have moved to the window because I wanted to see whether my friend had arrived, because I was influenced by a previous hypnosis session, because I was ordered to, or because I was attracted by light. There is nothing in the signals explored above (efference copy, perceptual reafferences) that can tell the agent whether he has done the action deliberately or not, through his own choice or under the influence of external cues. The signals in question have the function of helping an agent recognise whether he did move or whether the environment changed independently, but they don't help him recognize whether *he has been made to move* at a "deeper" causal level. The second sense in which an action is voluntary consists in attributing to the agent's mental states an *active*, versus an *instrumental* role, in initiating the action.

By an *active role*, is meant that the event that caused the bodily movement is some set of conscious beliefs or of motivations (it should be kept in mind, however, that most internal states and events, such as storing information, or producing an emotional response, are *not* themselves voluntary processes). This is the kind of account of action favored by folk psychology, and it is to it (in this sense) that the word *voluntary* usually applies. In the terms of Wolfgang Prinz (1997) summarizing critically this standard view, "Voluntary actions appear in the world because actors have a certain *goal* that they *want* to attain, and because they believe that this goal can be attained with the help of a certain action". (1997, 157).

By an *instrumental role*, is meant that the relevant causal event for triggering the bodily movement in the agent is some subpersonal mechanism harnessed to a particular property in the environment, making the agent directly susceptible to be set to act by that property. For example, an agent sees an object of interest, which triggers an arm extension for taking hold of this object. Similarly, a red traffic light automatically leads the driver to apply the brake.

This sense of a voluntary action allows specifying our question : are minimal actions voluntary in this B-sense ? Given that, as we saw, they are by definition not caused by an intention with a specific content, but rather by specific cues in a given context, they obviously *fail to be voluntary in sense B*. Although they feel like something done by the self, in sense A of a voluntary movement, they are not caused by any particular occurrent, conscious mental state. There is nothing in the awareness which a subject gains of his own "minimal" action which allows him to realize that "he was *not* in charge". No particular feature in the subject's experience can inform him whether he himself did what he did, or whether he was instead moved to do it by some subpersonal mechanism resonating to some feature out in the world. It is all the more striking that the absence of any distinctive phenomenological property for "active" versus "instrumental" agency does not leave, so to speak, a *gap* in the subject's experience. Each time he *moves* voluntarily, the agent seems disposed to believe that he thereby *acts* voluntarily. Agents engaging in minimal actions, as Wolfgang Prinz observes "do not do what they want, but they want what they do" (Prinz 1997b, 155, cf. Reader vol. II).

Most authors agree that the causation of an action is not located at the agent's level, but at the level of motivational and epistemic states. A bolder claim, defended among others by Prinz (Reader, vol. II) and Proust (1996), takes the agent considered as a person as a late - and maybe optional - construct in human development, with no primary causal role in action, minimal or full-fledged. In contrast, the very *belief* that one is a person carries a lot of causal potential. People learn first how to act, then strive for being identified and recognized over time by their group fellows as temporally extended responsible persons. This strife leads them to performing actions they would not have been motivated to do otherwise.

This analysis of persons in terms of beliefs should not lead us however to the view that an individual human being has no internal states, no goals and no desires of his own. Indeed in a system with no such states, no construction of selves could ever occur. What we should rather say is that the motivational and epistemic states have a causal power by themselves, independently of the particular belief in an overwhelming entity or person who would "own" these states and, so to speak, pilot herself with one part of herself.

The reasons why it would be mistaken to claim that all human actions are minimal is that, once we acknowledge that stored information does shape current behavior, then the capacity of mental states to jointly cause an action has to be admitted. Of course,

the stored information has not been *generated* by the agent. The context contained the relevant kind of information which the system was able to extract, categorize and store. Desires were largely generated as a consequence of social interactions and phylogenetic programming. They did not arise as a consequence of purely endogenous events. This granted, a distinction should be made between those episodes where an action is carefully thought over and planned, and those where an agent acts with no conscious, verbally expressible motive. Even if we question the ontological significance of persons in causation of actions, it seems implausible to jump to the claim that all kinds of action are "arational".

We saw above that discovering whether oneself moved, or whether the perceived environment moved, is a matter of perception. Considering now the sense B of a voluntary action, one might agree that knowing whether one acted voluntarily is a matter of inference; as emphasized by D. Wegner, the agent relies on the phenomenology of active movement to *infer* that he is also the conscious causal source of his own acts. As the well-known experiments of Nisbett and Wilson (1977) have shown, subjects tend to explain retrospectively their actions by invoking the standard reasons other people give for doing them, even when they in fact acted under some external pressure. In other words, they tend to mistake minimal actions for full-fledged ones: as we saw above, a minimal action is precisely not defined by its having an internal, mental cause. So, whereas minimal actions fall by definition in the realm of behaviors that imply a voluntary movement, they fail to imply the property of being voluntary in this deeper sense of having a long-term internal state as their cause.

Let us come back to our former example. Only in the first case ("I have moved to the window because I wanted to see whether my friend had arrived") is the agent acting voluntarily, in the sense that she has a specific reason for going to the window. If she goes to the window because she is influenced by a previous hypnosis session, she is certainly not acting voluntarily: she has no reason of her own to make this move; she is furthermore unaware that her doing so satisfies the intention of someone else (the hypnotist). This contrast warrants making a B distinction between "acting voluntarily" and "acting involuntarily". Some people may move voluntarily while acting in an involuntary way. They are just caused to act by some external agent or physical stimulus.

We must add an important caveat, however, when it comes to applying this distinction to specific cases reported in the first-person mode. Granted that the subject cannot introspectively discriminate whether her reason to act is what *really* moved her to act, it may well be that the action was not voluntary after all, in the sense that the agent was moved by some contextual cue rather than by her own desire to see her friend. *Her belief might rationalize her behavior while failing to have caused it.*

A further important notion is involved in this contrast; in case I go to the window to see whether my friend arrived, things had been different, e.g. had there been snipers around, I *would not* have moved to the window. In this kind of case, I can change my plan flexibly; in the hypnosis case, or in the command case, I cannot. So the contrast between voluntary and involuntary actions might involve in a number of cases a distinction between actions whose development (i.e. initiation and monitoring) is more or less automatic and those whose development is more or less sensitive to context. (The "more-or-less" clause is more plausible than a strict division between two categories. I can pace around in the room absent-mindedly, but still adjust to the context, and refrain doing so when it would involve unwanted consequences. Still automatic drivers often miss the correct freeway exit).

Still the B-way of understanding the contrast between voluntary and involuntary actions does not capture an additional, important aspect of voluntariness, having to do with the fact that an agent may -- on line or retrospectively -- *assent or dissent* to performing an action while performing it. A subject may have moved voluntarily, and have been caused to move by some long-term disposition of hers (thus acting voluntarily in the sense B above), while being unable to resist the force of her drive for performing the action. Her action should then count as C-involuntary. Let us see why.

C - Voluntariness as a second-order property of actions

Let us assume that, even when an agent's behavior is actually controlled by the environment, she represents her action as driven by internal motives. Generally speaking, an agent may or not, while acting, feel a dissonance between the desire or

motivation that triggered her action and some higher order desire. This dissonance may be more or less intense, according to the kind of action performed (compare : "I should not have killed him" with "I should not have smashed the vase"). Such a dissonance can be understood as a functional property, independent from the notion that the agent would be "free" to act otherwise. This dissonance implies a third level of analysis of a voluntary action.

Let us make sure that the subject matter of free will does not interfere with the present debate. Freedom to act as one wishes is an issue sometimes confused with the disposition to agree or not with one's first order motivations. What does it mean to say that one acts "freely" ? According to the so-called "principle of alternate possibilities", a person can act freely only if she could have acted differently from what she did in a particular situation. But given the fact that all the physical states are what they are, the idea that one could have acted differently from what one did - or thought differently from what one did - is a contentious matter, one which should not be taken for granted in naturalistic approaches to action. Indeed Wolfgang Prinz has convincingly shown that the notion of free will cannot be incorporated in an empirical psychological theory (Prinz, 1997b, cf. Reader vol. II). Let us thus assume in this discussion that every single state, mental or physical, is the effect of a set of physical causes such that no other act was possible for a given agent in a given context. This assumption has no bearing on the distinction between the case where an agent simply acts as she does without further ado and the case where she is able to ask herself whether she is right to do what she does. This latter capacity, as Harry Frankfurt (1988) insists, is a purely representational capacity ; it is *not* an executive one : it is independent from the ability of the same agent to *resist successfully* her first order motives. For example, a drug addict craving for heroin may form the second order desire *not to desire to take heroin*, while being unable - now and later on - to resist the desire to take it.

This kind of case allows distinguishing two C-varieties of a less than voluntary action. We will examine shortly how this distinction applies to our minimal actions.

a - Non-voluntary action as an impulsive and irreflexive action.

A first kind of case consists in acting under the influence of internal or external causes, without being able to represent alternatives, as is the case for an impulsive and irreflexive action (Frankfurt, 1988). An action is *impulsive* when it is performed on the basis of some first-order desire, or some current external cue, without any inhibitory mechanism being applicable. An action is *irreflexive* when the agent is not able to entertain second-order desires : she cannot consider alternative courses of actions and order them in the light of her preferences, or she "does not care" for the kind of desires that guide her actions. Non-human animals and young children seem to belong to this category of impulsive and irreflexive agents : they act on the basis of the stronger will activated, without being able to form a concurrent representation of other courses of action nor to select among their desires the kinds they would rather promote. An impulsive action may thus be perceived as one's own (A-voluntary), and be seen as consonant to some desire one has, or as contextually adequate, (B-voluntary), without being voluntary in the sense of being considered in the light of some higher-order desire (C_A-voluntary).

It is interesting to observe that an action involuntary in this sense never appears involuntary to an irreflexive agent who performs it. By definition, an irreflexive agent *does not appreciate the fact* that she does not raise the question of her preferences concerning her first order motivations. Only an agent able to represent her first-order desires can raise this question, a fact with a high moral and legal relevance.

Many minimal actions are impulsive, and their motives fail by definition to be considered. It makes them also involuntary "by definition" in the C_A sense. But this sense of involuntary does not account for an important difference between minimal actions. Some are in consonance with the general behavior of the agent, some are not. Driving absent-mindedly and smashing vases are two very different kinds of minimal actions. The following variety of C-non voluntary actions should account for this crucial difference.

b- Non-voluntary action as a motivationally dissonant action

In the preceding case of an impulsive-irreflexive agent, non-voluntary actions are

necessarily part of a repertoire of exclusively C_a - non-voluntary actions : a non-human animal can never reach the level of a reflexive agent, because it lacks the relevant set of representational-executive abilities. Human minimal actions, on the other hand, seem to be part of a larger repertoire of actions having different kinds of causal background, and more or less susceptible of being integrated with the agent's second-order preferences. Some of the actions of an individual may be performed impulsively, some others inhibited or carefully planned. Among those actions which are planned, some are planned using only first-order desires (as when one plans to prepare a gourmet meal) , some involve a selection between the sets of desires to be made efficient (should I rather eat as I wish and get fat, or follow a diet and get slim ?)

In the case of minimal actions, although they are not the product of a conscious intention to act, they can differ with respect to the evaluation which will take place, by necessity, *after* the deed. An agent may perform a large variety of minimal actions under the control of some external cue or present emotion, while being able to reconsider her actions reflexively after she performed them (or at most at the time she performs them). In this late reflexive stage, the subject may or may not, in Frankfurt's words, "identify to" her first-order motivation. For example she may recognize that crashing the China vase out of anger involved a substantial loss, that losing her self-control had many unwanted consequences etc., and regret her [minimal] actions. Or she may on the contrary find herself doing something which she later finds useful and in agreement with what she should have done (like ducking when a projectile was moving towards her, avoiding a car crash in a split second, etc.).

The important point is that, in all these cases of minimal actions, the agent may have acted involuntarily, in the sense (B) that she did not have the intention to act in such a way, and find herself either pleased or discontent with what she did. In the latter case, her action can be said to be involuntary in the sense (Cb) of being not only under no direct control of her desires, but also *in complete discrepancy with her second-order desires*. Intuitively, this contrast is exemplified in analogy with the situations of two drug addicts described by Frankfurt. It is one thing to take heroin because one desires to take heroin and because one thinks that taking heroin is a behavior that one wishes to promote ; another thing to take heroin because one cannot refrain from doing so, although one desires to stop taking it. In the same way, one can perform a minimal action which one can later identify to. Had she had time to think about it, the agent would have done just what she did, as she herself explains. Ducking was the thing to do. Although she did not duck voluntarily, there is one sense in which she is pleased with what she did, and would do it deliberately had she a chance to find herself in the same context.

Now take the vase smasher. Had the agent been given a chance not to smash the vase, she would not have smashed it. The agent here dissents from what she did, because she dislikes this episode of violence; she does not want to be moved by anger to do something. A minimal action of this kind is involuntary in the sense C_b that the agent is made to act against her own preferences.

This third level of analysis of how a minimal action can be seen as voluntary may raise the following objection. Why take an action which the agent *could not have prevented to happen* as voluntary on the basis of counterfactual properties, of the type "Had I been able to think things over, I would have/would not have done it". Is not this kind of consideration *too late* to call the corresponding action a voluntary one ? Here is the line that an answer to the objection might take.

If one considers the case of a single minimal action, having this third level of analysis of voluntariness may indeed appear as a useless sophistication : too late is too late. Granted that actions are reproduced over time, however, the evaluations which an agent produces of her past minimal actions in the light of her second-order desires may in the future help her control impulses of the same type. The second-order desires and similar *evaluative states* are therefore not idle, even when they arise after the action considered ; they have an important role in shaping the future behavior of the same agent, helping her to act voluntarily, in sense C, when circumstances allow it. The importance of the evaluation is that, in the future, a minimal action of the same type may, through an efficient inhibition, be prevented to happen, or at least, become subject to an inhibitory learning process. Even if one rejects the folk-psychological standard account of the will, there is thus room for an interesting role of voluntary action in the evaluative sense of the term to characterize a behavior that is shaped over time by second-order desires.

References

- Armstrong, D.M. (1991). Intentionality, perception and causality : Reflections on John Searle's *Intentionality*, in E. LePore & R. Van Gulick (Eds.), *John Searle and his Critics*. Cambridge: Blackwell, 149-158.
- Bach, K. (1978). A representational theory of action, *Philosophical Studies*, 34, 361-379.
- Bratman, M.E., (1987), *Intentions, Plans, and Practical Reason*. Cambridge: Harvard University Press.
- Daprati, E., Franck, N., Georgieff, N., Proust, J., Pacherie, E., Dalery, J. & Jeannerod, M. (1998). Looking for the agent, an investigation into self-consciousness and consciousness of the action in schizophrenic patients, *Cognition*, 65, 71- 86.
- Davidson, D. (1980). *Essays on Actions and Events*. Oxford: Clarendon Press.
- Decety, J., Jeannerod, M., & Prablanc, C. (1989). The timing of mentally represented actions. *Behavioural Brain Research*, 34, 35-42.
- Fourneret, P. & Jeannerod, M. (1998). Limited Conscious monitoring of motor performance in normal subjects. *Neuropsychologia*, 36, 11, 1133-1140.
- Frankfurt, H. (1988). *The importance of what we care about*. Cambridge: Cambridge University Press.
- Goldman, A. (1970). *A theory of human action*. New York: Prentice Hall.
- Goodale, M.A., Pélisson, D. & Prablanc, C. (1986). Large adjustments in visually guided reaching do not depend on vision of the hand or perception of target displacement. *Nature*, 320, 748-750.
- Hursthouse, R. (1991). Arational actions. *Journal of Philosophy*, 88, 2, 57-68.
- Jeannerod, M., & Fournet, P. (1998). "Etre agent ou être agi? Sur les critères d'auto-attribution d'une action". In H. Grivois & J. Proust (Eds.), *Subjectivité et conscience d'agir, approches clinique et cognitive de la psychose*. Paris: Presses Universitaires de France, 75-97.
- Jeannerod, M. (1994). The representing brain, neural correlates of motor intention and imagery. *Behavioral and Brain Sciences*, 17, 187-245.
- Jeannerod, M. (1997). *The cognitive neuroscience of action*. Oxford: Basil Blackwell.
- Lhermitte, F., Pillon, B., Serdaru, M. (1986). Human autonomy and the frontal lobes. Part I: Imitation and Utilization Behavior. *Annals of Neurology*, 19, 4, 326-334.
- Libet, B. (1985). Unconscious cerebral initiative and the role of conscious will in voluntary action. *Behavioral Brain Science*, 6, 529-566.
- Mele, A. R. (1992). *Springs of Action*. Oxford: Oxford University Press.
- Nisbett, R.E. & Wilson, T.D. (1977). Telling more than we can know : verbal reports on mental processes. *Psychological Review*, 84, 3, 1977, 231-259.
- Prinz, W. (1987). Ideo-motor action. In H. Heuer & A.F. Sanders (Eds.), *Perspectives on Perception and Action*, Hillsdale, NJ: Lawrence Erlbaum, 47- 76.
- Prinz, W. (1990). A common coding approach to Perception and Action. In O. Neumann & W. Prinz (Eds.), *Relationships between Perception and Action*. Berlin: Springer-Verlag, 167-201.
- Prinz, W. (1997a). Perception and Action planning, *European Journal of Cognitive psychology*, 9, 2, 129-154.
- Prinz, W. (1997b). Explaining voluntary action : the role of mental content. In M. Carrier & P.K. Machamer, (Eds.), *Mindscales, Philosophy, Science and the Mind*. Konstanz : Universitätsverlag, Pittsburgh: Pittsburgh University Press, 153-175.
- Proust, J. (1996). Identité personnelle et pathologie de l'action. In I. Joseph & J. Proust (Eds.), *La Folie dans la Place, Pathologies de l'interaction, Raisons Pratiques*, 7, 155-176.
- Proust, J. (1997). *Comment l'esprit vient aux bêtes*. Paris : Gallimard.
- Proust, J. (1999a). Indexes for action. In *Revue Internationale de Philosophie, Neurosciences*, 3, 321-345.
- Proust, J., (1999b). Experience, action and theory of mind. In *Developmental Science*, 2, 3, 286-287.
- Proust, J. (2000). Awareness of Agency : Three Levels of Analysis. In T. Metzinger (Ed.), *The Neural Correlates of Consciousness*. Cambridge: MIT Press.
- Proust, J. (to appear₁). Imitation et agentivité. In J. Nadel & J. Decety (Eds.), *Imitation, représentations motrices et intentionnalité*, Paris: Presses Universitaires de France.
- Proust, J. (to appear₂). Perceiving intentions. In J. Roessler & N. Eilan (Eds.), *Agency and self-awareness: issues in philosophy and psychology*. Oxford: Oxford University Press.
- Rosenblueth, A., Wiener, N. & Bigelow J. (1943). Behavior, Purpose and teleology. *Philosophy of Science*, 10, 18-24.
- Roth, J., Prinz, W. & Maasen, S.. (2000). *Voluntary action, A Reader*. Texts collected to prepare the Conference on Voluntary Action, Delmenhorst, Germany.
- Searle, J.R. (1983). *Intentionality, an Essay in the Philosophy of Mind*. Cambridge: Cambridge University Press.
- Searle, J.R. (1991). Reference and intentionality. In E. LePore & R. Van Gulick (Eds.), *John Searle and his Critics*, Oxford: Blackwell, 227-241.
- Sommerhoff, G. (1990). *Life, Brain and Consciousness, New Perceptions through Targeted Systems Analysis*, North Holland, Amsterdam: Elsevier.
- Sperry, R.W. (1950). Neural basis of the spontaneous optokinetic response produced by visual inversion. *Journal of Comparative and Physiological Psychology*, 43, 482-489.
- Von Holst, E., & Mittelstaedt, H. (1950). Das Refferenzprinzip, Wechselwirkungen zwischen Zentralnervensystem und Peripherie. *Naturwissenschaften*, 37, 464-476.