Indexes for action

In *Revue Internationale de Philosophie,*
Volume Neurosciences, 1999, 3, 321-345.

Joëlle Proust

CNRS, CREA
Ecole Polytechnique, Paris

Should philosophy revise its own theories in the light of new scientific facts ? Although philosophers do not all agree on a positive response, naturalist philosophers consider that no good philosophy can be built upon outdated science. This is not to say that philosophy can be reduced to science. Whereas scientists are dealing with facts, philosophers have primarily *conceptual* duties - with the correlative obligations for keeping an eye on possible worlds, theoretical alternatives and questions left unexplored by empirical investigation. Philosophy occupies a notoriously difficult position : while it should extend its conceptual considerations to non-realised possibilities, it should still be attentive to facts, in the sense that a conceptual analysis incompatible with empirical data cannot claim to be adequate.

One of the clearest cases of an outdated philosophical claim is offered by substance dualism, which considers mind and body as two distinct entities with different causal capacities, a view that neuroscience has made increasingly difficult to sustain. Other dualism-related philosophical claims have been shown to rely on untenable scientific tenets; the notion that consciousness can be understood as some central function enjoying some unified perspective on the mental processing (Dennett, 1993), and preparing in a chronologically anterior session all the relevants steps in decision and action (Libet, 1985). In these and similar cases, it appears in retrospect that modern and contemporary philosophy have unduly taken folk psychology at face value.

The present paper will concentrate on one of the domains in which contemporary neuroscience seriously upsets the picture offered jointly by folk psychology, traditional philosophy and even not so distant experimental psychology, namely action, its representation and its relationships with other mental functions. Until recently, philosophers of mind considered that the motor part in action did not intrinsically belong to the mental domain. In the dualist tradition, the faculty of understanding defined the mind, whereas the faculty of moving deliberately, i.e. of executing the decisions taken at a cognitive level, was considered a lower-level bodily function[1]. Executing was not seen as requiring per se anything like specific representations of the environment and of the body; more exactly, motor behavior appeared to rely on perceptual and conceptual processing that was independently effected upstream. Classical psychology also assumed that motor output would simply use perceptual information collected, so to speak, in a more or less speculative, "all-purpose" way. One of the corollaries of this claim was that particular motor outputs had no privileged relation with given perceptual inputs; being non-cognitive, motor coding was considered to be neutral on the issue of sensory representations; in other words, there was in this perspective no such thing as classes of movements being more readily accessible given certain classes of sensory input, over and above the associations between stimuli and responses that learning allows establishing.

Philosophy of mind, as well as philosophy of action, have been deeply influenced by the view according to which cognition is mainly "input related". All major philosophers of perception analyzed perception as having to do with objects and events in the outer world. The view of perception as a detached form of sensory access to the external world is currently challenged by a set of neuro-physiological findings. Ungerleider & Mishkin's 1982 theory of the "two cortical visual systems" suggested that two types of neurons are involved in vision. While an occipito-temporal system is involved in object-recognition and semantic association of visual characteristics, a parieto-frontal system is specializing in visuomotor control; according to this analysis, visual neurons in the dorsal stream mostly respond to "where" questions, while those in the ventral stream mostly answer "what" questions. This theory was later refined in different ways[2]. Of particular interest is the observation that the dorsal system includes in turn two distinct ways of coding that are directly correlated with representing action-related properties (Jeannerod, 1997). Spatial properties that are extracted through perception are either "intrinsic" or "extrinsic" (or relational) object properties. The first ones are examplified by shape and size; they determine the characteristics of the motion that are involved in

---

[1] See in this respect Missa's rebuttal of Bergson's view of action as a non-representational cerebral function in Missa (1997).

[2]In particular, Zeki (1993) insisted that the relevant distinction was not so much between object related properties, on the one hand, and spatial properties, on the other, as between two different ways of extracting forms ; the ventral one extracting form and color, and the dorsal one extracting form in motion. But this analysis was still mostly "vision oriented". Even though it did emphasize the role of the dorsal system in visuomotor transformations, it was still grounded in the notion that visual representation is organized though modalities having to do with visual coding.

grasping an object. The second, like location in space (represented in egocentric coordinates), or motion dynamics, are characteristics involved in *reaching* towards an object.

These results suggest that philosophers of perception may have been biased in several ways in analyzing perception. First, perception is far from being a passive registering of external properties, objects and events, contrary to a common philosophical view; it seems to involve some kind of precategorization of perceptual content; to perceive is to represent something *as being* somewhere, and *as affording* some kind of movement. Such a tacit categorical knowledge is displayed in the interactions of the perceiver with the world, for example through the various anticipations that attentional mechanisms provide. Secondly, perception involves a complex interplay between egocentric and allocentric data. Philosophers absorbed into externalist views on perceptual content may have to refine their approach in order to encompass the full range of perceptual operations.

One further important consequence of this and other neurophysiological findings is that philosophers now have new tasks regarding the analysis of action itself. Standard philosophical definitions of action explain it in terms of some reason to act which causes a relevant bodily movement. In this vein, Davidson (1980) holds that a proattitude relative to some state of affairs, in combination with the relevant belief that such and such a movement would bring about the desired condition, produces causally the instrumental movement. Thus the reason to act is both a *justification* of an action, and the *cause* of the corresponding physical behavior. What is questionable in this analysis is *not* that reasons to act should be instrumental in producing the external movement; although Wittgensteinian philosophers insist that any conflation between causation and rationality is a philosophical confusion of the "grammatical" variety, many philosophers take the causal role of reasons to be a necessary precondition for an intentional action; how good would be the Aristotelian practical syllogism if it was denied any neuronal realization? The problem is rather that this type of analysis does not address at all the question of how this realization could proceed, and even seems to defy any plausible solution. Once the action schema is taken to include propositional attitude contents, on the one hand, and a physical movement, i.e. a motor event, on the other hand, it is not clear how the gap between world-directed intentional contents and motor execution could ever be bridged, which in turn opens new concerns about the so-called problem of deviant causal chains[3].

[3]A second problem, to which I will return later, is that such an analysis ignores a wide category of actions that are not performed "for a reason", on the basis of a set of appropriate beliefs and desires. This second problem will be dealt with below.

The problem of deviant causal chains consists in the fact that the intentional content of an action can be brought about causally by the agent in some unintentional way. One of the classical examples is the case of a nephew who drives to his uncle's home in order to kill him with a gun, and out of nervosity runs over a pedestrian who happens to be his uncle[4]. There is actually a causal link between the reason to act and the physical behavior leading to the outcome, but that causal link does not qualify as the right one, for it is not part of the intended relation between action and outcome. One cannot, by definition, try to perform an action by using some non-controllable kind of movement. Nervosity is clearly failing to qualify as a rational means to reach the intended end. This kind of external factor of deviance can clearly be matched by internal factors, where some inappropriate cerebral condition releases a motor program that should have been inhibited, given the present motivational states of the subject, and that leads to a movement inappropriately caused, and therefore not properly intentional. An example of a deviant internal causal factor is offered by the case of a person X who wants to tell the truth to Y about some fact, and tells it when intoxicated by alcohol instead of doing it in a controlled manner.

Although Davidson has his own reasons for doubting that the problem of deviant causal chains can be solved at all,[5] a philosopher may want to resist them and imagine various ways to address this difficulty. One consists in asking the agent's confirmation that such and such a causal chain was (or failed to be) part of his intentions. But as Ogien (1997) notes rightly, one cannot rely on *the agent*'s appreciation of his action - as to whether it was rightly caused - without recognizing the incompleteness of the above analysis in terms of mental states. Indeed one of the interesting features of Davidson's analysis of action is that the agent is not himself causal in an action, which allows a finer-grained analysis of action causation.

It should further be stressed that, when one speaks of mental causation, the agent's report on which mental states are actually causal may be unilluminating, insofar as the agent is in no way an infaillible judge about why he did what he did. Empirical data tend to show that subjects have no priviledged access to the mental states that are causal in their own decision to act[6]. This should not lead to deny that something like the mental-states causation picture is right, but simply to acknowledge that the reasons which the agent gives for his own actions may not always be the ones that were instrumental in producing the action.

---

[4] Cf. Searle (1983), 82, and Chisholm (1966), 37.

[5] His reasons have to do with his refusal of psycho-physical laws ; cf. Davidson (1980), 80.

[6] See Nisbett & Wilson (1977).

Another perhaps more promising way, consists in providing an analysis of intentional action that bridges the gap between intention to act and execution of the relevant action, in a way that clarifies the causal contribution of the intention in the actual movement. This kind of analysis has been attempted both by neuroscientists and by philosophers trying to bring neurophysiology to bear on the problem of action causation.[7]

The general idea is that the crucial factor that explains how an action can be normally caused by some prior or occurrent intention is the existence of some unique set of pragmatic rules -also called motor representations - that build up and control the formation and the execution of intentions to act. Clearly, the very existence of a jointly representational *and* executive medium *would* make the question of deviant causal chains more tractable. The data unfortunately do not warrant such a uniform view on action. There are at least two distincts types of action codes[8]. One describes the action in allocentric, relatively abstract and general terms ; it can be consciously accessed and memorized, and can be also expressed verbally; it is the semantic dimension of action, that involves long-term associative memory. The other code represents the action in an egocentric way. It is usually rapidly forgotten, and if not entirely implicit, it often defies precise verbal expression. At first blush, neurophysiological two-tiered construction of action has to solve a problem similar to the problem of deviant causation which philosophical analyses have raised : how could potentially incommensurate representational formats be connected in order to produce a unique goal-directed behavior ? In the present paper, I will present three elements that both offer a solution of these two closely related problems and qualify as constituents of the very concept of an action: 1) a movement must be regulated by *feedback* to qualify as an action ; 2) some *historical-teleological* connection must be present to achieve the connection between pragmatic and semantic aspects of the action ; 3) finally, an action involves crucially the extended use of *deictic reference*.

1- Feedback

The concept of *feedback* in the philosophy of action has been used against the traditional view according to which an agent accomplishes successfully an action only if he is able beforehand to represent in detail all the relevant aspects of its interaction with the environment[9]. This traditional view presents the difficulty that

[7] See Bach, (1978), Jeannerod (1994), Livet (1994, 1996, 1997), Pacherie (1997a), (1997b), (1998a), (1998b), Proust (1996, 1998).
[8] See Jeannerod (1994) et (1997).
[9]See Livet (1994)

the subject is supposed to be able to act thanks to some retrospective influence of the goal on the action under development, i.e. by way of some so-called "final cause". The first attempt at showing that the concept of a goal-oriented behavior can be analysed without having to use any final causation can be found in Rosenblueth,Wiener and Bigelow's classical 1943 paper. They suggest that a goal-oriented behavior can be realized mechanically (i.e. by a purely physical system) if only the system can receive and use a negative feedback "from the goal". The notion of feedback generally requires that a system turns part of its output into an input; one speaks of*negative* feedback when the signals involving the target constrain the output in order to reduce the error between the target location and the current trajectory.

Presence of feedback is often considered a central piece in the definition of action[10]. Now there are different ways in which this kind of feedback may be involved in action. These different views do not respond in the same way to the problem of deviant causal chains. An inflow version is exemplified by Pierre Livet's initial 1994 model. An explanation that radically eliminates finalism from action theory should according to Livet put all the burden of the intentional dynamics on feedback. In such a theory, the target, i.e. the final state of the system is the only element that needs to be represented (in egocentric coordinates). The movement develops from a weak initial constraint to a sequence of corrections - by negative feedback - that bring the organism to the target location. In this kind of explanation, feedback is temporally coincident with intentional behavior, while looking ahead - or anticipating the result - is confined to the activation of the representation of a final state.

Still this representation does not work as a final cause. In such a model, the causation of action involves both goal representation and the inflow sensory feedback about local dynamics relative to goal. As Pierre Livet emphasizes, it is only when the course of action is corrected that the agent *discovers his own intention* (Livet, 1997, 346). In this way, the chances for an action causal chain to be deviant are reduced by the fact that the actual course is recognized as constituting the corresponding intention. If the action is defined only by its actual corrective procedure, then there should be no discrepancy between intentional and efficient causation. A failure in feedback operation should activate an interruption of the course of action, and a recognition that its conditions of satisfaction failed to be met.

However it may be objected that in this analysis the gap between the goal realization and the sequence of corrections is not bridged, but only displaced. For if the only information represented beforehand is the target event, one can think of

[10]See among others Sommerhoff (1990), Livet (1994) et (1997), Proust (1997).

deviant circumstances that allow the motor intention to develop in the course of a corrective procedure, with an effect that is mistakenly taken to be the realization of the goal. As far as the goal representation is not *instrumental* in the correction procedure, the two intentions may well fail to converge, given the strong drives that may be exerted on the system by interfering stimuli. Indeed the very coincidence between intended effect and observed effect in a particular action seems in this analysis to be a miracle. The possibility that the action reaches its success conditions through non-intentional, causally deviant means is left fully open[11].

What is needed is a closer link between action-monitoring and goal-representation. Another view on the role of feedback turns this link into a constitutive relation in which the goal controls the feedback at each step. This view consists in making feedback *subordinate to* the representation of a central command allowing predictions in sensory reafferences. It is articulated in Arbib's model for a "coordinated control program", containing perceptual and motor schemas, as well as feedback loops between sensory input and schema assemblage. (Arbib, 1990). It has also been developed in connection with neurophysiological data in a series of works by Jeannerod and his collaborators. (Jeannerod, 1994, Jeannerod, 1997).

In contradistinction with the preceding theory, feedback is not only provided by the peripheral senses, but is also internally fed under the anticipatory monitoring of a central command, the feedforward signal. Internal information about expected input at each stage of action is directly provided to the relevant sensory structures at a central level. A comparison can thus be effected on line between the steps towards the desired goal and the action being executed. Neurophysiological evidence brings support to the hypothesis that such a functional organization could allow a much quicker and more flexible capacity for adapted action than some version of purely external feedback. The concepts of corollary discharge[12] and of efferent copy[13] both express the same crucial idea: the brain sends signals that allow keeping track of its own commands and uses them to *compare* outflow information (commands issued and sensory predictions) and inflow information (sensory reafferences)[14].

Feedback is thus now understood in a very different way. A system first generates predictions on the basis of the command issued, (on the basis of earlier

---

[11] For an interesting comment on and reevaluation of Livet (1994), see Livet (1997).
[12] Sperry (1950)
[13] Von Holst & Mittelstaedt (1950)
[14] See Jeannerod (1997) for a detailed analysis.

actions and their stored cue-effect pairs), then compares its predictions with actual reafferences. Feedback is for this reason considered "internal". This model explains both how feedback can be used without delaying performance, and how motor trajectories are designed by feedforward very early in the course of a simple action, such as a grasping. It has been shown, for example, that the hand is preshaped, early in the movement, in a way adapted to the shape of the target. (Jeannerod & Biguer, 1982).

This decisive notion of a command-based control allows explaining in part how causation can proceed smoothly from preparation to execution. Nevertheless it remains to be explained how motor control per se can be linked to the semantics of action, which is the crucial element for deviant causal chains. As we saw earlier, deviant causation threatens when there is no nomological link between intending to effect P and effecting P. We just bridged the gap between intending P and doing the motor act towards P. The problem now is to bridge the remaining gap between doing the motor act towards P and realizing P. How is it at all possible that this particular movement is done as part of that particular action ? Given that there exists apparently an infinite number of possible movements compatible with a particular action, there seems to be no causal link that could warrant the unicity of the action considered semantically, i.e. through its intentional content, and the action considered motorically, i.e. through its pragmatic content.

2 - Schema selection

Philosophers have frequently observed that an action can be given a variety of descriptions that are all consonant with the intentional content of the action. Feinberg called it "the accordion effect". John Searle illustrates this effect through the example of Gavrilo Princip's murder of Archduke Franz Ferdinand in Sarajevo: " Of Princip we say that he: (1) pulled the trigger, (2) fired the gun, (3) shot the Archduke, (4) struck a blow against Austria, (5) avenged Serbia"[15]. In other words, one can distinguish the basic action, i.e. the bodily movement effected with a given intention, from its results. Among those, some are actually part of the *semantic content* of the action, such as items (3-4-5) in the accordion above; other results of the action are part of its *pragmatic content,* such as pulling the trigger. The accordion effect reflects the fact that a basic action, defined as a bodily movement executed in virtue of a certain set of beliefs and desires, has many different results ; among them, holds a link which can be causal (such as 1-2-3) or conceptual (such as 3-4).

---

[15]Searle (1983), 98.

The intentional results are those that are in a means-end relationship, whereas the non-intentional ones are results that are no part of the conditions of satisfaction of the action.

Philosophers have entertained different views on how many actions are performed in the list above. While Goldman[16] considers that the list represents an "action tree", i.e. a complex of actions, each one being individuated by its relevant property (e.g. effecting a trigger-pulling vs effecting a gun-firing), Davidson defines the action as the bodily movement that is directly caused by the corresponding mental states. According to Davidson, all the above characterizations of Gavrilo Princip's murdering of the Archduke refer to the same action defined as a bodily event, although they offer various properties to describe it. "We never do more than moving our bodies : the rest is up to nature."[17]

One way of rephrasing the problem of causal deviance at this stage of our reflection is to try to clarify what exactly constrains the relationship between pragmatic and semantic contents of an action. John Searle[18] solves the related problem of how to assemble the prior intention with the intention in action by invoking a common feature of them both : they are *self-referential*, which helps the causal link to develop in the intended manner from prior intention to bodily movement. The prior intention is self-referential in the sense that its content is that "itself be causally efficacious in producing the intended outcome", and the intention in action is self-referential in the sense that it expresses the requirement that itself should be causal in the bodily movement. According to Searle, the following clause expresses the content of the total intention to act :

> (I) that this prior intention *causes* [this intention in action *causes* this bodily movement].

In Searle's view, this content of the intentional content of an action must be fully grasped by an agent if it is to count as this agent's intention; further, it constitutes the corresponding *conscious* experience of acting. These two claims have raised a number of objections, among which I will select the following three[19]. 1) Should the *awareness to act intentionally* depend on the grasp of embedded self-referentiality of intentions? 2) Is it plausible to require from this grasp that it capture the *causative* relation between the two kinds of intentions, thereby excluding non linguistic animals and human infants - beings who presumably lack concepts of

16 Goldman, (1970).
17 Davidson, (1980), 59.
18 Searle, (1983), pp. 93 sq.
19 See Bach, (1978), Armstrong, (1991), Pacherie, (1998b).

intentions, of causality, or of both - from the range of intentional action? 3) Finally, should the conscious attempt to act determine the range of actions in general ? Many actions seem to be performed by an agent without any kind of *conscious* intentional content. If this is so, how are intentions in action to be distinguished from bodily behaviors of the deviant kind ?

I want to suggest that a negative answer should be offered to questions 2 and 3, while taking a positive stance towards the first: I will defend the view that self-referentiality is indeed an essential component of awareness of acting intentionally (as well as of acting *tout court*). Although such a view is by no means original (see Pacherie, 1998), it has been associated with the claim that action has to be *consciously* initiated or performed to count as intentional. I will on the contrary recognize with other authors (Bach, 1978, Frankfurt, 1988) that some actions, which Bach calls "minimal", are performed unthinkingly, automatically, routinely, and even in ways that are controlled by the environment and not by the agent's conscious or unconscious will. Such a recognition seems to be consonant with neuroscientific and neuropsychological approaches of action, as numerous actions in humans and non-humans are found to be triggered not by an antecedent intention, nor by what Searle calls an intention in action, but by the simple *perception* of some relevant stimulus[20].

If our concept of action is to include minimal actions, our concept of intention must itself be correspondingly generalized by allowing "minimal intentions". Any motivating property, be it a reason to act or some external stimulus poised to "drive behavior" could in principle constitute an intention to act. As this use of the term may be found counterintuitive, I will simply replace it, in claim I above, with the corresponding type of representation at work:

I': that this semantic representation of action A *causes* [this pragmatic representation of action *causes* this bodily movement].

We are thus left with two difficulties. 1) How are we to understand the role that the term "causes" plays in determining the intentional content of an action in a way that does not require that the intentional content *depend on* grasping the concept of a cause ? 2) How is self-referentiality able to glue together, so to speak, both types of action representations, i.e. the semantic and the pragmatic ?

---

[20] James, (1890), Frith, (1992), Berthoz, (1997).

The first element of a solution of how an intentional action can include causality without that causal link between the semantic and the pragmatic representations being explicitly represented in the intentional content resides in the historical-teleological aspect of action. A classical explanation for the association between goal representation and movement execution has been offered by the "action-effect principle", developed by James and Lotze, and is now inspiring, if I am not mistaken, in one form or another, the work of many contemporary theorists of action, among whom Berthoz, Jeannerod[21] and Prinz[22]. According to this principle, an action is acquired, planned, and controlled in terms of its effects. As we saw earlier, feedback from the effects can be stored centrally and used in determining further courses of action. As the distal effects coincide with the external outcome of the action, it is through the semantic properties of the action (that it has such and such a consequence on the perceived external world) that an action is represented and stored. Lower-level pragmatic representations are thus subsequently activated by the "higher-order" goal.

Jeannerod offers as an illustration of this regimentation of pragmatic concepts under semantically defined action concepts with an experiment by Marteniuk et al. (1987), in which subjects have to reach for a disc in two conditions : either they have to fit it into a box ; or they have to throw it[23]. What is found is that the subjects have a slower reach movement and a longer phase of deceleration in the first condition, which can only be explained by the characteristics of the *ulterior* part of the action. The pragmatic representation seems thus to be under the selective control of some final goal, as predicted by the action-effect principle[24]. Some neural cells in the premotor cortex seem indeed specialized in coding a movement for a certain goal, whether performed by the agent or by another person[25]. This principle can be applied not only to simple sequences of action, but to more complex action-plans that introduce a succession of hierarchical constrains on nested goals of a more and more concretely determined variety.

The action-effect principle is certainly of heuristic value, even though it may appear too crude in the formulation given above: the action-effect principle as stated is indeed clearly applicable only in the case when an action can be repeated from the same initial to the same final positions. But such a situation is certainly the

---

[21] A difference between their theories that needs not concern us here bears upon the role of internal vs external feedback in the effects of an action.
[22] See Prinz (1997) & Hommel (1997).
[23] Jeannerod (1997), 126.
[24] Fuster (1995) shows the role of two sets of overlapping prefrontal networks representing respectively the sensory cue and the motor response (cf. 179 sq).
[25] Di Pellegrino et al., 1992.

exception, given that the agent moves his body and changes his dynamic properties, and that the target event may also have variable kinematic and other physical and motivational properties. The principle of action-effect has to be completed by a specification of how the optimal trajectories given the target location and the biomechanical constraints can be learnt. Such a specification involves a simulation of an action and a comparison between object position coded in extrinsic coordinates and hand location coded in intrinsic coordinates. Principles of optimization should be used by the system to select the most adequate pragmatic representation for the represented goal. Whatever refinements can be brought to the action-effect principle, the latter allows explaining in non-final terms how, in Prinz's words, "the idea of a movement could bring about the realization of that movement"[26].

For philosophers, the interest of the empirical solution described in the "action-effect principle" terminology is that it can be redescribed in a more abstract way which reveals the *teleological* structure of this theory. To be able to do X in order to get some desired result P it must be the case that :

(1) in the past, tokenings of X produced randomly P in P'-motivated contexts ,

(2) the very fact that former tokenings of X produced P caused a new token of X to be activated in new P'-motivated contexts.

In the above analysis, P' is the motivational cue that is present when X is first effected, with the result of obtaining a motivationally adequate outcome P. For example, let us suppose that some hunger state obtains, and that some movement X is executed in a random way in presence of food, satisfying the need of the hungry animal. A side result of this initial sequence is that in similar motivational contexts (hunger), the pragmatic representation of mouth opening will be more likely to be activated, causing in turn the corresponding movement.

This first step allows explaining how beliefs and desires, or in the generalized view, semantic representations, can *become* efficient in *causing* a bodily movement. The answer is that the corresponding pragmatic representation has been initially learnt as a means to reach an end, which allows later to solve the so-called "inverse problem" of which means has to be realized for an end to be reached. Schema selection results from the above association between motivational context, and pragmatic concept activation.

It also explains some important fact about intentions and semantic representations of actions, i.e. their *normative* character. An intention to act does not only entertain *causal,* but also *normative* relations relations with its effects, in the sense that some

[26]Prinz, (1994), 218.

outcome is *supposed to happen* as part of the very etiological causation of the relevant bodily movement. Not only is this outcome the one "desired" or "willed" by the agent that performs the action ; it is also an outcome that is part of the history of the present action. As Millikan (1993) has convincingly shown, normativity supervenes on evolutionary explanations of the presence of a token. Once some replication mechanism insures that some element can be copied when it has been helpful to the organism's survival, the copied element can be evaluated for the contribution that it is supposed to make to the welfare of the organism in virtue of its belonging to its particular type. In the case of action, replication is the result of learning ; obviously, the tokening of the representation of the whole action sequence depends functionally on some *prior* success rate of the corresponding type of representation. Normativity is a pervasive characteristic of action : not only is a bodily movement *caused* by prior learning ; the system *anticipates* the reafferences, and, in Anokhin's terms, "approves" the correct development of action on the basis of expected sensory feedback.[27]

The action-effect principle thus explains the causal mechanism in virtue of which a semantic and a pragmatic representation can become associated when two neural configurations become co-activated by learning. Once such an association is established, it allows the agent to know at least roughly how to proceed. Therefore, the agent does not have to represent such a causal mechanism as part of his intentional content for the action to unfold in a goal directed way[28].

We are now ready to confront the second problem raised above. How is self-referentiality able to glue together, so to speak, both types of action representations, i.e. the semantic and the pragmatic ? Searle's analysis above suggested that self-referentiality was a crucial element in the solution. But how can

[27] See Berthoz (1997), 17 sqq.

[28] Pacherie (1998) offers an alternative analysis of pragmatic representations that has interesting features in common with the present action-effect principle. She suggests that pragmatic representations treat objects as causally indexical, (like "easy to handle with one hand"). The term of "causal indexicality", borrowed from Campbell (1993), refers to the fact that specific objects are treated as affordances, i.e. as goals-for-me, by a motivated organism. Causal indexicals as understood by Campbell can thus be viewed as egocentrically determined functional entitities, i.e. correlates of possible goal-directed behaviors, whose functional relevance is learnt through the organism's antecedent activity. They allow, just as the action-effect principle does, to acknowledge practical grasp of the object's value for action without involving any conceptual understanding of what grounds the success of the action. Although the action-effect principle fails to match the linguistic elegance of causal indexicality, the latter fails to explain how an organism acquires the practical knowledge of how to use this or that object given some current needs.

such a self-referentiality be realized in non-linguistic animals or in minimal actions ? More exactly, how can it be realized in terms that do not involve, circularly, the perspective of a conscious agent ?

3 - Monitoring of action, deicting binding and self-referentiality.

One tempting way of explaining why a self-referential analysis of the intentional content of an action solves the problem of deviant causation consists in the following scenario, close to Searle's, of what it is like to act. The agent should, while acting, in substance think : "this goal [or semantic representation of the action] of mine, G, is currently being achieved by my A-ing", where "my A-ing" refers to some particular current activation of a pragmatic representation. The co-instanciation of the same anchoring (*my* willing G, *my* doing A) seems sufficient to warrant an adequate causal connection between both types of representations for the same action.

However common-sensical, this analysis raises several difficulties. First, as Pacherie (1998) correctly observes, the above analysis has to be completed with a demonstration that *this token* of willing is causally involved in *this token* of doing. It could well happen that the agent lost track of time, and mixes up two different successive attempts at performing an action. This is the problem of the temporal binding of the two representational components of an action.

Second, it can very well be that introspection simply cannot give us the leverage on what the right link between the two representations are. We saw above that the agent's grasp of the causal link between the two representational aspects of his action plays no crucial role in the existence of that causal link. The agent has, in ordinary circumstances, a unitary experience of doing one thing -- illuminating the room; not the experience of doing one thing *by* doing another -- illuminating the room in virtue of pressing a switch. His learning of the causal link may be entirely implicit, and develop in the absence of any concept of the relevant properties involved in the successful performance of his action, whether in the external world or in his own body. In a similar way, the agent could simply miss the crucial elements involved in his action representations.

Still one could insist that the agent knows that *he does this now in order to reach G* because he entertains some conscious mental state that guides his action from planning to final execution. This is the strategy that Pacherie (1997) and (1998) uses to forbid wayward causation of action. In her model, the relevant mental state for causal self-referentiality is offered by *motor imagery:*

> "The motor image provides the organism with an awareness of what is intended and with a conscious grasp of his body as generator of acting forces". (22).

A motor image indeed concentrates the two kinds of information accessible in a self-centered way, whose conjunction seems to provide a solution to our deviant causation problem : one is a determined, conscious representation of the final state to reach; the other is what the agent currently feels as his own dynamical progress towards that state. Binding over time and binding means with ends is achieved in this theory by making self-representations explicit at these two crucial representational levels.

However elegant and well-thought, this theory raises several difficulties. One forceful objection against the view that conscious access to one's own performance can deflect any deviance in causal links consists in the fact that the agent may have *illusions* of agency[29]. He may believe wrongly that a certain motor image is what allows him to act (for example, as dance teachers know well, he may have in mind the motor image of, say, a pas de bourrée, and execute in fact something quite different). He may also believe wrongly that he acts at all ; he may be dreaming, or, as anosognosic patients do, report having performed actions that he is, in fact, totally unable to perform.

We have two other reasons to refrain from using the agent's introspective access to his own intentions to act. First, we need independently to understand what it is for an agent to be a person, and can hope to do so on the basis of specific properties of the organism's mental states. The concept of a conscious state is thus to be developed without presupposing the concept of a conscious agent. Although mental imagery could, at least in principle, be approached as a mental state in abstraction from the agent who entertains it, the explanatory role that imagery has in the development of one particular action calls forth introspection, and thus agent-level properties.

Second, as neuroscientists and psychologists well know, it is always in principle possible to substitute a functional property for an introspective one[30]. This observation becomes all the more relevant when our analysis aims at including "minimal" actions, performed without any conscious imagery : a functional explanation should be more general than a "phenomenological" one.

These difficulties invite an analysis of self-referentiality of action representations that leaves aside the conscious properties of these representations. Could not we

---

[29] Cf. Bach (1978).

[30] Leaving aside philosophical considerations of the "inverted spectrum type". See Block (1978) on the limits of functionalism about qualia.

find functional properties capable of anchoring the components of an action representation to one place, one temporal episode, and one particular way of achieving it, i.e. identify *binding* properties ? A plausible candidate would have to satisfy the demands of *action monitoring*. For any action to be performed up to its end, from instigation to completion, some monitoring has to take place, that maintains activated the representation of the goal - this particular goal - across a changing context, guides the feedback in the way indicated above and brings the process to a halt when the goal-event is reached. As a consequence, every step of the action is performed *under the monitoring of the very plan to which it belongs*. Now the question is : how does the brain achieve this monitoring in an ever-changing external context ?

Recent considerations on *deictic binding* could offer interesting perspectives for answering this question. A deictic representation is a representation whose content depends on the context of a particular informational capture. Classical examples of deictic representations are demonstrative expressions : *this* object, *here*, *that* man. Ostensions also function as deictic variables, whose referent is the object indicated by the ostension-token. Recent work by Ballard et al. (1997) suggests that the brain might use deictics to represent actions. Self-referentiality of action or action components, independently specified by philosophers in response to the problem of deviant causal chains[31], would thus have a litteral application at the level of the cerebral codes used for action monitoring.

Reasoning in abstract functional terms, there are three types of deictic representations that need to be involved in any action to make it a single functional entity. *Perception* first locates *where* or *on which* object to apply the action program. *Working memory* then maintains this rule as a *context* for further steps in the action. Finally, the *motor performance* relies on the variables delivered by the working memory to ensure proper internal feedback, and on the deictics delivered by current perception to get the correct external feedback. Both sources will be used to *terminate* the action.

Crucial in the whole process is the acquisition of referents in what could be called a "deictic program". Vision for example can be modeled as a "deictic pointing device" : the fixation point (on a certain object) constitutes the center of an external frame of reference, in which the action is supposed to take place (either now or later). Grasping an object normally involves looking at it and locating it in the outer world in allocentric coordinates. Other senses can also be used as deicting pointers: haptic manipulation, audition, (and even olfaction in many animal species) can help

---

[31] See in particular Searle (1983), Searle (1991) and Pacherie (1998b).

determine where the action should take place[32]. In all cases, the perceived object is instrumental for having the relevant pointer activated; the action will thus preserve the pointer's content as part of its conditions of satisfaction, be it a particular object , location or motor routine.

In this model, binding action components over time depends on the successive activation of the relevant pointers. Some of them will be activated all along, keeping track dynamically of a particular location, when it comes to spatial tasks[33]. Some of them will be activated sequentially, according to prestored programs. Cerebrally, the crucial information seems to be delivered by the basal ganglia; they code delayed reward through dopaminergic neurons, and thus allow step by step program sequencing (Jeannerod, 1997). Some neurons seem to be specialized in indicating that a previous step is terminated, others that some next step is to be performed[34]. The caudal neostriatum would provide for short-term visual indexes[35]. All these atomic, deictic operations would be under the control of the prefrontal cortex. Those prefrontal neurons are in effect responsible for delayed responses (Fuster, 1973)[36] and have been shown to be involved in a form of mnemonic processing (Fuhanashi et al., 1989). The term of *working memory* is used to refer to all kinds of domain-specific processes "for the proper utilization of acquired knowledge" (Goldman-Rakic, 1991). The frontal cortex is thus taken to be the cerebral structure that allows working memory to be loaded with the representation of the task and to constrain execution in a task-relevant way.

Conclusion

I considered three ways in which the connection between semantic and pragmatic representations of a single action can be tightened up. A first move consisted in making the feedback process, i.e. the dynamics of the relationship between both representational components, a central element in the definition of an action. A second step brought in the action-effect principle, emphasizing the teleological relation of each pragmatic representation type with its external effects. A final step consisted in elucidating the constitutive character of demonstrative reference for

---

[32] Action can also be mental, and will necessitate in that case "neural pointing devices": cf. attention. See Ballard et al., (1997).

[33] The concept of dynamic indexicality was first explored by the philosopher Gareth Evans ; see Evans (1985), ch. 10 ; cf. also Pacherie (1998).

[34] See Kermadi & Joseph, (1995), quoted in Jeannerod (1997), p. 157.

[35] Caan et al., (1984), quoted in Ballard et al., (1997), p. 738.

[36] Cf. also Fuster (1995)

the contents of working memory states. Did we succeed in alleviating the charge that deviant causal chains could disrupt not only our empirical understanding of an action, but also our capacity to individuate actions, and to define what an action is ?

A philosopher could complain that our present model tends instead to open Pandora's box of deviant causal chains, for there now seems to be as many chances that causation goes wayward as there are distributed codes jointly activated to represent and control a particular action. After all each particular deictic representation *could* be causally connected to the wrong percept or to the wrong internal state.

There are three reasons not to take this worry too seriously. One is that we can - at least in theory - rely on our teleological account for what a normal effect is, thanks to the action-effect principle. Therefore we may find out what the normal causative process should be in a family of cases belonging to a same learning process. This should take care of the classes of cases of the nephew-killing-uncle variety. The second is that the connection of a deictic element with its referent is a matter of how the world is, and not a matter of how the organism has been wired, or of how it takes the world to be. Indeed one of the interesting aspects of an indexical view of the structure of action is that the external world - with its past and present properties - plays quite a constitutive role in the selection of action programs as well as in the actual course of the action, with all its near-misses, anticipations and backfirings. This view thus offers an externalist understanding of action, with the consequence of dispensing us with the "Cartesian-style" philosophical difficulties[37]. Third, the very number of representational levels that neuronal analysis has to take into account should discourage a simplified view on causation. If one causal link is failing, others should be able to cope with the situation : correcting misperceptions, reestablishing temporal sequence, trying again etc. This is what action is all about.

Acknowledgement : The present article considerably benefited from stimulating discussions with Jean Decety, Bernhard Hommel, Marc Jeannerod, Pierre Livet, Elisabeth Pacherie and Wolfgang Prinz, to whom I express my gratitude. Mistakes remain mine.

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.

[37] I borrow this expression from McDowell's (1986) article.

Berthoz, A., (1997), *Le sens du mouvement,* Paris, Odile Jacob.
Block, N., (1978), Troubles with functionalism, in W. Savage (ed.), *Perception and Cognition, Minnesota Studies in the Philosophy of Science, 9*, 261-325.
Chisholm, R..M., (1966), Freedom and action, in K. Lehrer (ed.), *Freedom and Determinism,* , New York, Random House.
Arbib, M. A., (1990), Programs, schemas and neural networks for control of hand movements : beyond the RS Frameword, in M. Jeannerod (ed.), 111-138.
Bach, K., (1978), A representational theory of action, *Philosophical Studies,* 34, 361-379.
Ballard, D.H., Hayhoe, M.M., Pook, P.K. & Rao, R.P.N., (1997), Deictic codes for the embodiment of cognition, *Behavioral and Brain Sciences,*
Berthoz, A., (1997), *Le sens du mouvement,* Paris, Odile Jacob.
Caan, W., Perrett D.I. & Rolls, E.T., (1984), Responses of striatal neurons in the behaving monkey, , *Brain Research,* , 290, 53-65.
Campbell, J., (1993), The role of physical objects in spatial thinking, in N. Eilan, R. McCarthy & B. Brewer (eds.), *Spatial Representation,* Oxford, Blackwell, 65-95.
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.* Vol. 65, pp. 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.
Dennett, D., (1991), *Consciousness explained,* Boston, Little, Brown & Co.
Fuhanashi, S., Bruce, C.J., & Godman-Rakic, P., (1989), Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex, *Journal of Neurophysiology,* 61, 331-349.
Fuster, J.M., (1973), Unit activity in prefrontal cortex during delayed response performance,. Neuronal correlates of transient memory, *Journal of Neurophysiology,* 36, 61-78.
Fuster, J.M., (1995), *Memory in the Cerebral Cortex,* Cambridge, MIT Press.
Goldman, A., (1970), *A theory of human action,* New York, Prentice Hall.
Grivois, H. & Proust, J., (dirs.), (1998), *Subjectivité et conscience d'agir, approches clinique et cognitive de la psychose,* Paris, Presses Universitaires de France, 1998.
Hommel, B., (1997), Toward an action-concept model of stimulus-response compatibility, in B. Hommel & W. Prinz (eds.), *Theoretical Issues in stimulus-reponse compatibility,* Elsevier Science, 281-318.
Hornsby, J., (1980), *Actions* , London, Routledge and Kegan Paul.
Jeannerod, M., & Fourneret, P., (1998), "Etre agent ou être agi? Sur les critères d'auto-attribution d'une action", in H. Grivois & J. Proust (dirs.), 75-97.
Jeannerod, M., (1993), Intention, représentation, action, *Revue Internationale de Psychopathologie,* 10, 167-191.
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.
Jeannerod, M., (ed.), (1990), Motor representation and control, *Attention and Performance* XIII, Hillsdale, Lawrence Erlbaum.
Jeannerod, M., & Biguer, B., (1982), Visuomotor mechanisms in reaching within extrapersonal space, in D. Ingle, M.A. Goodale & R? MAnsfield (eds.), *Advances in the analysis of visual behavior,* Cambridge, MIT Press, 387-409.

Kermadi, I., & Joseph, J.P., (1995), Activity in the caudate nucleus of monkey during spatial sequencing, *Journal of Neurophysiology,* 74, 911-933.
Kornblum, S., Hasbroucq, T & Osman, A., (1990), Dimensional overlap : cognitive basis for stimulus-response compatibility - a model and taxonomy, *Psychological Review, 97, 2,* 253-270.
Libet, B., (1985), Unconscious cerebral initiative and the role of conscious will in voluntary action, *Behavioral Brain Science*, 6, 529-566.
Livet, P. (1994), *La communauté virtuelle,* Combas, l'Eclat.
Livet, P. (1997), Modèles de la motricité et théorie de l'action, in J.L. Petit, (ed.), (1997), 341-361.
McDowell, J., (1986), Singular thought and the extent of inner space, in Ph. Pettit & J. McDowell (eds.), *Subject, thought and context,* Oxford, Clarendon Press, 137-168.
Millikan, R., (1993), *White Queen Psychology and other essays for Alice,* Cambridge, MIT Press.
Milner, A.D. & Goodale, M.A., (1995), *The visual brain in action,* Oxford, Oxford University Press.
Missa, J.N., (1997), La théorie bergsonienne du "Cerveau, organe de l'action", in J.L. Petit, (ed.), 99-110.
Nisbett, R.E. & Wilson, T.D., (1977), Telling more than we can know : verbal reports on mental processes, *Psychological Review*, vol. 84, 3, 1977, 231-259.
Ogien, R., (1997), Pourquoi il est si difficile de "naturaliser" l'action, in J.L. Petit, (ed.), 27-37.
Pacherie, E., (1997a), Troubles de l'agentivité et troubles de la conscience de soi, in J.L. Petit, (ed.), 363-385.
Pacherie, E., (1997b), "Motor images, self-consciousness, and autism", J. Russell (ed.), *Executive Dysfunctions in Autism*, Oxford, Oxford University Press, 215-255.
Pacherie, E., (1998a), Représentations motrices, imitation et théorie de l'esprit, in H. Grivois & J. Proust (dirs.), 207-248.
Pacherie, E., (1998b), The content of intentions, *Rapports et documents du CREA,* CREA, Paris, N° 9724.
Petit, J.L. (ed.), (1997), *Les neurosciences et la philosophie de l'action,* Paris, Vrin.
Prinz, W., (1987), Ideo-motor action, in H. Heuer & A.F. Sanders (eds.), *Perspectives on Perception and Action,* 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., (1997), Perception and Action planning, *European Journal of Cognitive psychology, 9,2, 129-154.*
Proust, J., (1996), L'expérience du mouvement, Perception et motricité, *Actes du séminaire Ecologie des Transports Urbains*, 109, 75-103.
Proust, J., (1997), *Comment l'esprit vient aux bêtes*, Paris, Gallimard.
Rosenblueth, A., Wiener, N. & Bigelow J., (1943), Behavior, Purpose and teleology, *Philosophy of Science, 10,* 18-24.
Searle, J.R., (1983), *Intentionality, an Essay in the Philosophy of Mind,* Cambridge, Cambridge.
Searle, J.R., (1991), Reference and intentionality, in E. LePore & R. Van Gulick (eds.), *John Searle and his Critics,* Cambridge, Blackwell, 227-241.
Sommerhoff, G., (1990), *Life, Brain and Consciousness, New Perceptions through Targeted Systems Analysis,* North Holland, Amsterdam, Elsevier.

Sperry, R.W., Neural basis of the spontaneous optokinetic response produced by visual inversion, *Journal of Comparative and Physiological Psychology,* 43, 482-489.
Ungerleider, L., & Mishkin, M., (1982), Two cortical visual systems, in D.J. Ingle, M.A. Goodale & R.J.W. Mansfield, (eds.), *Analysis of visual behavior,* Cambridge, MIT Press, 549-586.
Von Holst, E., & Mittelstaedt, H., (1950), Das Reafferenzprinzip, Wechselwirkungen zwischen Zentralnervensystem und Peripherie, *Naturwissenschaften, 37:* 464-476.
Wolpert, D.M., Ghahramani, Jordan, M.I., (1995), An internal model for sensorimotor integration, *Science*, 269, 29 september 1995, 1880-1882.
Zeki, S., (1993), *A vision of the brain,* Oxford, Blackwell.
