Are empirical arguments acceptable in philosophical analyses of the mind?

Joëlle Proust

HAL Id: ijn_00000068

https://jeannicod.ccsd.cnrs.fr/ijn_00000068

Submitted on 26 Jun 2002

Are empirical arguments acceptable in philosophical analyses of the mind?

J. Proust
Institut Jean-Nicod (CNRS, EHESS)
jproust@ehess.fr

Is it appropriate for philosophy to borrow from science theoretical concepts or to invoke experimental evidence in order to do its explanatory work? One central argument for a positive answer is that rational explanation in general has a holistic character. Such holism applies both to the inferential procedures taken to be valid and to the beliefs taken to be true. Discussions of the notion of reflective equilibrium have developed during the last decades these two central themes. 1) The principle of *reflective equilibrium* as described by Nelson Goodman (Goodman, 1965) shows that the process through which our forms of inference are justified is virtuously circular: particular deductive inferences are justified by valid general rules that are themselves justified by the fact that they allow valid inferences. 2) What holds for inference can be generalized to the whole explanatory process. John Rawls (1974) offered a wider picture of reflective equilibrium, including not only the inferential part of the justificatory process, but the substantial part as well: our inferences constrain and are constrained by all the semantic, epistemological, metaphysical and psychological knowledge available. What holds for justification in general seems to hold *a fortiori* for philosophical arguments in epistemology, independently of their authors' specific claims and doctrines.

Other arguments however have been offered in favour of the autonomy of philosophical explanations relative to science. The main one is that philosophical analyses in general should be *a priori*, i.e. established independently of empirical investigation. The notion of a prioricity here at work is aimed at reflecting the particular stance that philosophy should adopt towards experience: not as something simply given, which might be offered an empirical description, but as a something to be known and justified in a rational discourse, involving a "logical space of reasons" rather than "a logical space of nature".[1]

There is clearly a tension between these two approaches of rationality, a tension that may be eased with a more thorough discussion of the kind of empiricism to be endorsed, as can be found in the recent literature. Let us assume for now that the question whether empirical considerations may be used in general philosophy is settled one way or another; the question of their use in the *philosophy of mind* would still remain unsolved. For even if one accepts that experts are in better position to know what the concepts we borrow from them mean, one may resist deferring to experts when it comes to one's own mental experience. In the theory of collective rationality sketched above, experts are a central ingredient in the justification of most of our judgements. For a state of reflexive equilibrium can only be reached if everyone recognizes that *mathesis universalis* is a socially distributed matter. To be able to reach knowledge, one needs to justify one's beliefs against the backdrop of a coherent set of beliefs as

---

[1] The latter expression is meant by McDowell (1994) to offer a symmetrical phrase to Sellars' "logical space of reasons".

available in any potentially relevant domain of thought and inquiry. Obviously a social division of labour is needed here: an individual thinker has to defer to the experts of specialized domains both for offering theoretically justified definitions as well as for providing lawful regularities and testable explanations. Indeed, as Grice (1965) and Putnam (1975) suggested, speakers do defer to experts to provide the precise definitions of terms they use. Scientific research itself works in part on the assumption that the terms now used may be later redefined as part of the explanation which the researchers are presently striving to provide[2].

Now although the basic structure of explanation should be the same in all the parts of philosophy, the philosophy of mind offers an additional source of difficulty. A single "unsophisticated" individual, with no particular scientific training in theoretical psychology and sociology, seems indeed entitled to justify thoughts about herself. She is thus in a position to explain her own conduct, to plan and make rational decisions even though she is lacking any theoretical knowledge about psychological and sociological facts. More important, still, she can do so even though she does not know that such a theoretical knowledge is available at all. It seems that the principle of rejecting such an unsophisticated individual from the realm of rational agents would introduce a dangerous segregation among thinkers, and would also put intolerable constraints on successful communication between speakers of different cultures.

It would also invite scepticism about self-knowledge. For once the instability of the knowledge presently available in the neurosciences is recognized, a thinker, however "sophisticated" she may be, would also have to accept a principled deference in time. A rational thinker would thus have to accept that she may not be able to justify presently her own mental concept use, or the way she derives her decisions from her past experiences. There is a new tension here, that does not concern the possible gap between experience as *given* (in the scientific scheme) and experience as *justified* (in the philosophical scheme), but rather between *everyday understanding of oneself* and *theoretical explanations about the self.* This tension will be the focus of the present article.

Two contrasting positions, expressed by Jennifer Hornsby and Georges Rey, will help us sketch the two opposing positions.

Jennifer Hornsby: " We ought not to assume at the outset that the basis of our everyday understanding of one another is susceptible of correction and refinement by experts in some specialist field where empirical considerations of some non-commonsensical kind can be brought to bear". (1997, 3-4)

Georges Rey: " Surely it is obvious "at the outset" with regard to any explanatorily interesting project that people who have studied the phenomena more systematically than is ordinarily possible will be in a position to correct and refine the basis of our ordinary thought." (2001)

Both authors are concerned here with the question how autonomous philosophical reasoning about the mind should be from a scientific view of the world - including theoretical psychology and the neurosciences. Or, in other words, they are discussing how close and exclusive is the connection between a philosophical approach of intention, action and self-understanding with a commonsense, folk-psychological approach of the mental. This question has been addressed mostly by authors who fight

---

[2] See also Rey (1997).

or defend naturalism. My first section however will attempt to show that the question is not to be simply identified with a general stance on the naturalistic program. It is not to be assimilated with the question of how philosophy generally relates to science, although clearly this general problem has consequences on the present discussion. It has to do with the acceptability of empirical arguments in the philosophy of mind, which is clearly an epistemological topic, dealing with justification of claims of a certain kind. I will then turn to the philosophical claims that are generally brought to the fore to justify the autonomy thesis relative to the philosophy of mind. I will discuss how independent these claims can be, with their respective bearing on the autonomy thesis, and will discuss their validity.

1 - Empirical considerations and the naturalistic program

To be clearer about the relation of our present question with a general position on the naturalistic program, we need to determine at the outset what is understood under this term. Dewey defined "naturalism" (in Baldwin's *Dictionary of philosophy and psychology*) as a methodological project for philosophy: it is "the theory that the whole of the universe or of experience may be accounted for *by a method like that of the physical sciences,* and with recourse only to the current conceptions of physical and natural science; more specifically, that mental and moral processes may be reduced to the terms or categories of the natural sciences".

Now all the work in this definition is done by the notion of "a method like that of the physical sciences". How can a philosophical argument *resemble* one in the physical sciences? By being couched in causal terms, whose relata do not differ from the entities usually recognized in the natural sciences - in particular in a non-normative way? But if such was the case, a philosophical argument would become indistinguishable from a scientific argument, and therefore philosophy would dissolve into specific scientific inquiries.

The notion of likeness may however be understood in a methodologically less constraining way, meaning that a naturalistic approach rejects "spiritual or transcendental arguments" in the terms of Dewey. In this sense, naturalism is a metaphysical claim, according to which only material entities matter causally; naturalism boils down to materialism, i.e. the monist claim that the substance supporting causal relations in nature is one and the same in all the diverse phenomena, whether mental or physical.

It is generally believed that most modern naturalistic philosophers belong to the methodological rather than to the metaphysical kind of materialism[3]. In other words, most of them do not have as a goal to show that the mind can be reduced to a material or a physical system, in the strong sense that mental regularities should be explained entirely in non-mental terms. They are content with a view in which although mental states are identical with physical states, mental laws cannot be directly expressed at a physical-causal level.

But we will not examine any longer the various positions that are currently debated on the question of which kind of physicalism can be found acceptable to account for the mental causal power. For the question we are dealing with *cuts across* specific views one may have towards the metaphysics of the mind. The difficulty noted above has to do with how justified someone is to refer to one's mental states in one's own terms, irrespective of the teachings of science on the metaphysics of the mental. The specific way in which a theory of the mental is dealing with physicalism is irrelevant to it. A second reason why naturalistic considerations are of no help is that while naturalism offers a *general* approach in which philosophy is not insulated from the sciences, but on a par with them, -- depending on a common lore of knowledge and rational constraints --, it does not offer any *specific* recommendation on how philosophy is able to play its own score within the general symphony of the sciences. The methodological recommendation of being attuned to a scientific view of the world does not bring with it a specific set of rules for using science, or mimicking scientific method. Surely it won't do to simply clothe one's own reflections in a scientific garb, as Hume presenting his

---

3   See for ex. Gary Hatfield's *The Natural and the Normative*, (p. 17).

study of the mind under Newton's authority. Naturalism does not recommend that philosophy becomes a kind of loose registration of facts: surely naturalism does not invite philosophers to making scientific results accessible to the many. A third reason is that the very idea of following a naturalistic method may not be the relevant kind of question that should be addressed in *every* philosophical inquiry. Using or not empirical arguments might be decided differently according to the specific subject matter one deals with. For example, someone studying the philosophy of biology might want to apply, say, the concept of function, in a way compatible with the facts stated in contemporary evolution theory and genetics; whereas someone studying moral value or formal demonstration might be justified, at least to a certain extent, in ignoring psychological and sociological facts about morality or reasoning. The deep reason is that natural science does not cover all there is to think about, even in a world exhaustively composed of matter.

A way of generalizing the three arguments above is to claim that the question of the validity of empirical considerations is also raised within non-naturalistic frameworks. Even if one defends the view that thought is inherently normative and irreducible to scientific description, one will need to appeal to "normative intuitions", concerning what seems right or wrong to individual subjects, and even to "normative facts", as Kant does when reflecting on the conditions of possibility of true knowledge as ascertained in the various existing sciences. Here again, one needs to establish how a subject would be in a position to reach rational evaluations about her own beliefs and plans if she does not master the kind of normative facts and intuitions that are best exemplified in the mature sciences.

For all these reasons, I will detach the question of the acceptability of empirical arguments in the philosophy of mind, from the stance one should take on naturalism as a view on the architecture of knowledge.

2 - The autonomy of philosophy in the philosophy of mind.

The autonomy of philosophy, in this context, is the claim that an appropriate answer to the central questions of the philosophy of mind should be delivered by philosophical investigation and argument alone, without relying on data or arguments from the empirical sciences. More specifically, the view is that, unlike natural kind concepts, such as chemical elements or pathological syndromes, whose definitions have to be established by scientific investigations, mental concepts are essentially defined through their ordinary usage, i.e. through common-sense psychology. Of course, presented in this way, the view is expressed in a question-begging way. For why should common-sense psychology be more akin to philosophical investigation and argument than theoretical psychology? Why, in other words, would philosophy be more "at home" in common-sense psychology than in theoretical psychology or in neuroscience? And why should common-sense arguments and reasoning belong to the realm of *a priori* reasoning?

A response to this question has been offered by various philosophers of this century and the last, including Wittgenstein, Anscombe, McDowell, and many others. The central role of the view which relies on common sense to study the mind philosophically can be established through a distinction between *two kinds of intelligibility*. McDowell presents the latter in the following way. A first kind of

explanations is one "in which things are made intelligible by being revealed to be, or to approximate to being, as they rationally ought to be". A second kind is that "in which one makes things intelligible by representing their coming into being as a particular instance of how things generally tend to happen"[4]. The idea then is that, whereas philosophy brings intelligibility to mental facts by showing how these facts explain rationally an agent's action through *her own* access to psychological and external facts, science explains mental facts by showing *how they are generated* and *"tend to happen"*. Obviously, the distinction between two kinds of intelligibility is tightly related to the contrast between a normative and a descriptive perspective on things. For in the first case the analysis essentially comes up with reasons to act, whereas in the second kind of case, things are represented as they happen in certain causal contexts. In the case of the philosophy of mind, it also is conceptually connected with the contrast between a first person and a third-person perspective on mental facts. Granted that common-sense psychology allows deploying the kind of normative intelligibility pinpointed by McDowell, whereas scientific psychology explains behavior in a purely causal-descriptive way, commonsense psychology is all there is to know to construct arguments in the philosophy of mind.

The distinction between two kinds of intelligibility is obviously not to be taken itself as a simple descriptive fact about knowledge. It depends on a more fundamental, overarching claim called the constitutive ideal of rationality, and it entertains conceptual relations with other claims about causality and self-knowledge, which we will try to explore.

A - The constitutive ideal of rationality

The notion of intelligibility in general which plays a fundamental role in McDowell's considerations in his 1985 paper, "Functionalism and anomalous monism", is borrowed from Davidson's central claim in "Mental Events"[5], about the holistic and rationally constrained character of a theory of mental attribution:

> "(..) When we use the concepts of belief, desire, and the rest, we must stand prepared, as the evidence accumulates, to adjust our theory in the light of considerations of overall cogency: the *constitutive ideal of rationality* partly controls each phase in the evolution of what must be an evolving theory". (p. 223)[6]

Many philosophers, among whom McDowell, base their appreciation of what functionalism can and cannot achieve, on the role played by what Davidson called the "constitutive ideal of rationality". We shall see that this claim also drives philosophers' stances on how scientific evidence can, or cannot, be incorporated to a philosophical analysis of the mental.

The idea of a constitutive ideal of rationality is that the understanding of other beings and of oneself is governed by principles that are articulated not by the laws of science but by rationality itself; these principles are not only used to understand others, they also make any interpretation possible. This last feature is what makes the ideal a "constitutive" one.

---

[4] McDowell,(1985).

[5] Davidson (1980), 207-227.

[6] see also *ibid*. p. 222 : "It is a feature of the mental that the attribution of mental phenomena must be responsible to the background of reasons, beliefs, and intentions of the individual".

Recognizing that the constitutive concept of rationality is a norm rather than a fact allows to appreciate, as McDowell does in his 1985 article, that beliefs and desires are involved in explanations "of a special sort", i.e. explanations putting "things" - thoughts, propositional attitude and actions - in perspective with what they ought to be. This indeed provides quite a different kind of explanation from one that predicts that, given P, Q tends to happen. If being thirsty, I take a beer in the fridge, this is not to be explained as a disposition to move in a particular way in a given context, but rather as a piece of behavior to be evaluated as rational, given the totality of my beliefs and desires at the time.

Even a theory that would model specific sets of deductive relations between beliefs that an agent may use to think and act, according to McDowell, would fail to credit rationality with a constitutive role, and therefore would not reach the level of normativity that is needed for the mental. Two features are of special importance:

- The kind of normativity involved is such that rationality is presented in an *entirely general way*. A norm is said to be "categorical"; in contrast, a set of domain-specific rules and inferences as specified in a functional approach would only offer *hypothetical* imperatives, i.e. *conditions* under which something qualifies as a system of beliefs.
-Mental normativity must be one in which rationality is presented in a way that captures the *applicability of forms to contents*. Common sense displays this kind of understanding, when it allows a subject to grasp that violating the norm of rationality automatically brings unintelligibility with it.[7]

B - The essential subjectivity of the mental

The notion of a constitutive ideal of rationality thus drives in turn an appeal to a subjective grasp, contrasted with an objective conformity, that will allow for spelling out how a norm can be espoused rather than simply followed. Thomas Nagel initially introduced his contrast between objective and subjective approaches of reality to emphasize the dynamics of objective knowledge and the limits of objectivity when it comes to understanding the self and other mental phenomena:

> " A succession of objective advances may take us to a new conception of reality that leaves the personal or merely human perspective further and further behind. But if what we want is to understand the whole world, we can't forget about those subjective starting points indefinitely; we and our personal perspectives belong to the world. One limit encountered by the pursuit of objectivity appears when it turns back on the self and tries to encompass subjectivity in its conception of the real. The recalcitrance of this material to objective understanding requires both a modification of the form of objectivity and recognition that it cannot by itself provide a complete picture of the world, or a complete stance toward it. (Nagel, 1986, 6)"

Nagel's problem however did not consist in emphasizing the *a priori* incompatibility between the approaches but rather in exploring the possibility of *integrating* them and clarifying the limits of such integration. In his perspective, "objectivity is a method of understanding". The method in question consists in stepping back from our own particular position in space and time, thus taking our own location as one unspecific position in the world, that we mean to characterize independently from ourselves, or "as it is in itself" (Nagel, 1986, 5). Nagel's suggestion consists in modifying the form of objectivity to accommodate subjective facts. The subjective form, which is for him the essence of raw, feels and intentional states can be only grasped by having them oneself, or else, by imagining that one has them. One can therefore think of oneself "from

---

[7] See McDowell, (1985), in McDowell, (1998), 330-1.

outside" while preserving the subjective way in which mental states are given to a subject, relying on "how far our subjective imagination can travel" (1986, 18).

In the philosophy of mind, the objective pull can however lead us astray: it is the case with what Nagel calls "objective blindness", a condition linked to ignoring the resources of imagination necessarily involved in any adequate understanding of minds and selves. This misleading attempt results in an "external theory" of the mental, i.e. a theory mimicking the natural sciences. Physics is clearly a science in which reality can be understood independently of a human perspective on the world. Philosophers draw on the case of physics to develop an approach of the mind that would similarly be detached from any particular perspective. Physicalism and functionalism are philosophical theories of the mental illustrating that strategy. Nagel feels that an unbearable separation within the self results from this vain effort of denying one's own subjectivity while studying mental properties.

This separation creates in turn the problem of "reintegrating" the external properties into living subjects:

> "One has to *be* the creature whom one has subjected to detached examination, and one has in one' entirety to *live* in the world that has been revealed to an extremely distilled fraction of oneself" (Nagel, 1986, 9)

Nagel does not explicitly indicate how his own method for grasping objectively subjective mental facts allows such a fusion between the creature and her mind. Part of the difficulty of Nagel's view is that it is expressed at a level of generality that does not bring much light to the way of conducting specific appropriate research on the mind.

In close agreement with Nagel, McDowell dismisses the view that "reality [about the mental] is objective, in the sense of being fully describable from no particular point of view" (McDowell, 336). There is "a natural intuition to the effect that the mental is both real and essentially subjective", he writes, now dissociating himself from Davidson's own position on this matter[8]. A fundamental mistake, according to McDowell, consists in "forcing the mental into an objective mould" (McDowell, 336).

McDowell understands "subjectivity" as covering not only phenomenal experience and its qualitative content, but also propositional attitudes. Two reasons are offered for treating propositional attitudes as subjective facts. First, entertaining them presupposes "comprehending the content of someone's outlook on the world". Second, distinguishing what does and what does not make sense involves "representing the idea that we might learn from others and thereby find them intelligible". What McDowell wants to show is that the subjective character of the notion of a limit to intelligibility has to do with the constitutive role of rationality. He explains the relationship between the two claims in the following way:

> "Achieving the kind of understanding for which rationality plays its constitutive role requires a sensitivity to the specific detail of the subjective stance of others, and an openness to learning from it, that is bound to be falsified if one supposes that explanations involving the constitutive ideal *work by locating their explananda in a structure specifiable from outside content"*. (McDowell, 1985, 337)(my italics)

---

2 For although McDowell says that Davidson "respects" this intuition, the latter does not dwell on it and even writes explicitly that "On the proposed test of the mental, the distinguishing feature of the mental is not that it is private, subjective, or immaterial, but that it exhibits what Brentano called intentionality", (*ibid.* 211).

This last claim provides the *closure* of the subjective experience and knowledge posited as the essence of the mental. It is in fact this central claim which *excludes an objective approach* as non-relevant in philosophy, and in particular forbids any illumination through scientific evidence. For without it, one might both concede that there is a subjective twist to propositional attitudes and to the openness of an individual to others, *and* defend the view that subjective properties can adequately be rephrased in equivalent objective terms (for example in sentences held true in response to given stimuli, or in dispositions to believe and desire).

Let us summarize claim 2. Objectivism consists in taking a third person point of view on the mind, i.e. adopting an external stance in which the mind is an object in the world among others, endowed with specific representational capacities. Using scientific results necessarily commits one to objectivism. Therefore another method, rejecting such commitments, should be used. This alternative method consists in approaching the mind in the subject's own terms, i.e. in common-sense idiom.

C - Mental causation and the common-sense view of the mental

Among the methodological consequences of the contrast between a subjective and an objective approach to the mind is the question of causality: should it be restricted to the objective realm, and shown irrelevant in the subjective domain of reasons? Or should a new causal picture, coherent with the contrast, be spelled out in more detail? In Nagel's perspective, there is only one kind of causation in the world, and the subjective realm does not provide ontology for a separate kind of causation. Even though the external perspective on the mind seems to dissolve the impression of freedom that subjects enjoy, no jumping out of individual minds outside time, space and physical causality is allowed: objectivity comes with an exclusivist view on causation. There is only one level of causation holding between events in the world.

> "The objective view *seems* to wipe out such autonomy [of the subjective intentional realm] because it admits only one kind of explanation of why something happened - causal explanation - and equates its absence with the absence of any explanation at all (Nagel, 1986, 115)

According to Nagel, this seeming however resists careful analysis. The search for an autonomous intentional explanation is seen as leading at best to a very limited kind of intelligibility[9]; it cannot explain "why I did what I did rather than the alternative that was causally open to me" (Nagel, 1986, 116). Furthermore, physico-bio-sociological causation is also part of our subjective experience: the objective pull maintains its grip on our sense of receiving passively from the world our capacities, our opportunities and even our own self :

> We remain, as pursuers of knowledge, creatures inside the world who have not created ourselves, and some of whose processes of thought have simply been given to us" (118) (..) The objectivity that seems to offer greater control also reveals the ultimate givenness of the self" (Nagel, 1986, 119)

Those philosophers who, like Nagel, recognize only one level of causation in the world, thus tend to recognize that empirical considerations may enrich the subjective stance through a process of self-distanciation and reappropriation. But another lesson on causation can be drawn from the existence of a subjective domain.

In contrast with Nagel, Jennifer Hornsby maintains that there is a distinctive level of causation in that domain that allows to completely account for actions "in the terms that we use as agents":

---

[9] At worst, it leads to incoherence.

> There is the thought, which Thomas Nagel has made especially vivid, that it is essential to our conceiving of ourselves as agents that we take our actions to be completely accounted for in the terms that we use as agents; the possibility of treating actions from the impersonal point of view would then subvert our ordinary conception of ourselves" (Hornsby, 1993, in Heile & Mele, 161).

The crucial notion in that argument is that of "conceiving ourselves as agents". For as we saw, this conception is in Nagel's perspective largely unsupported, and constitutes an irresistible article of faith for which no intelligible account can be provided.

Hornsby's particular way of arguing for this autonomous level of causal efficacy of the mental consists in maintaining that there is no overlap between internal, subjective causation, on the one hand, and external, objective causation, on the other hand. A defender of subjective view should maintain, according to her, that a direct causal foundation for intentional states and for agency can be found in what we might call a commonsense "world-and-psychology-view", characterisable independently from the world as described by physics, theoretical psychology and neuroscience. A typical question of an 'action explanation' is "Why did the agent do action A?" whereas a purely causal question is "Why was there an event of kind E?" The first question concerns a person, able to articulate her reasons in various propositional attitudes; the second has to do with an impersonal event in the objective causal order. For Hornsby, there is no common ground on which to deal with both questions. What we need only consider in explaining A's action is "a network of intelligible dependencies between the facts about what an agent thinks, what she wants, and what she does"[10].

She maintains that the causal-explanation view in the first sense (relevant for agency) is compatible with rejecting a view in which an action results causally from "discrete things interacting" (for ex. token states of belief and desire). (Hornsby, 1993, 167) Wanting and believing, in this view, are not events. Let us take for example the case of Peter wanting to boil some water. Peter's intentional states result in changing the world causally. The question how his intentional states are converted in physical forces is entirely external to philosophical concerns. There indeed is a gap between the subjective and the objective order, which Hornsby gladly acknowledges the gap. The agent may see the gap, but need not be concerned by it. He will notice that the gap is bridged when his action is successfully accomplished: "There is no need to invent an item to bridge the gap between Peter's states of mind and the event of his want's being satisfied: when his want was satisfied, the gap was bridged - he switched on the kettle".

Actions are initiatings of series of events (178) in this sense: there is something an agent intentionally does, there is some change in the world brought about by a person. In other words, an action is the event of a person's causing something, a movement of her own body, a letter written, a kettle of water being boiled, etc.[11] But what a person does, she insists, "is not an event, and it is therefore not a 'component of the flux of events in the world'". Or, as she also presents the view, "there are events that are not in the world "of nature"." (184). This makes a whole range of reductive questions

---

[10] Hornsby, (1993, 168)

[11] Davidson defends a similar view in his 1963 "Actions, reasons and causes". An intention to do A" does not refer to an entity, state, disposition or event" (in 1980, 8). Justifying and explaining an action "go hand in hand", but justifying depends on the explanatory role played by some intentional content (what Dadvidson calls "the primary reason"), and not conversely : the belief alone explains the action.

Davidson also claims that "there is something very odd in the idea that causal relations are empirical rather than logical" (14). His reason for this claim is that "a reason rationalizes an action only when the descriptions are appropriately fixed, and the appropriate descriptions are not logically independent" (14). The difference between Davidson's and Hornsby's treatments of action is that, for Davidson, causation can be accessed "impersonally", ie through mental states, whereas for Hornsby, only a person can have a causal power in acting. (see her 1993, p. 169).

irrelevant: At what stage in the neural chain shall we find an agent's action? "Our concepts of action may not contain the precision needed to determine an answer for it".[12]

The gap between the causal-explanatory approach in philosophy, on the one hand, and the causal-explanatory approach in psychology and the neurosciences is taken to be foreign to philosophical purpose by J. Hornsby. She writes:

> If the causal reality of belief and desire is just their causal-explanatory reality, then it need make no use of a further idea - of items inside people that we latch on to when we give action explanations" (1993, 168)

The upshot is that commonsense psychology is the proper medium for doing philosophy of mind; the ontological level on which causation normally supervenes being here bracketed, the only relevant level for philosophy of mind is the way people understand one another, i.e. commonsense psychology. (Hornsby prefers the term of *commonsense psychology* to the more usual expression of "*folk psychology*", because the latter implies that we are speaking of a kind of knowledge that specialists will reorganize around more efficient and testable concepts)[13].

3 - Discussing the three claims

Let us summarize the discussion at this point. We suggested that a stance on the value of empirical considerations in the philosophy of mind stems from three general claims concerning the study of the mental:
  1) The claim of the constitutive role of rationality
  2) The claim of the subjective essence of the mental
  3) The claim of the causal specificity of common-sense view on action and subjectivity.

The three claims that I summarized define a certain way of approaching mental concepts; my analysis emphasized their inter-relations. Let us observe again that the autonomy of the philosophy of mind can be defended in various ways. The stronger view presented above results from accepting the three claims, as Jennifer Hornsby does. The concept of a person, in Hornsby, is understood on the basis of a joint acceptance of the three claims. The person, rather than her mental states, causally determines her actions. She has a personal level access to her motives, and thus is able both to rationalize her actions and those of others.

It is possible however to defend only one or two of these claims while rejecting the remainder; significantly different stances on the mental would result, as well as different ways of dealing with the autonomy of philosophy (i.e. different ways of blocking, allowing or requiring empirical-scientific considerations into philosophy). McDowell, who espouses the first two claims, occupies an only slightly weaker position than Hornsby's does. Many non-reductionist philosophers of mind would agree with claim 1 (sometimes in a revised form) and with claim 2, while rejecting 3 (for example, Dennett, 1969). Davidson would only accept 1. Some hard-core naturalist philosophers would deny 1, 2 and 3.

Thus each particular combination of stances taken on the three claims implies a specific stance on empirical arguments in the philosophy of mind. The three claims listed above are obviously not jointly *necessary* to exclude recourse to scientific evidence from philosophical analysis of the mind (for there are certainly many other ways of concluding that empirical-scientific considerations are out of place in philosophy). As I just suggested, but did not have time to prove, they are separately *sufficient* to constrain the recourse to scientific neurology and psychology. I will concentrate on each of them.

---

[12] Hornsby, (1997, 67).

[13] Hornsby, (1997, 3-4)

Claim 1 and the autonomy of philosophy

It certainly is impossible in a few lines to discuss in detail the importance which McDowell attributes to the "constitutive ideal" of rationality, and how scientific reasoning, as a result, is taken to belong to a derived kind of rationality. We will concentrate on what explains that there should be such a constitutive link between truth and rationality, on the one hand, and intentional description and mental attribution on the other. Steven Stich (1990) has offered an explanation along these lines: "for a person's cognitive states to be intentionally characterizable, the states, the interactions among them, and their interactions with the environment must all be similar to our own".[14] Given the role which inferences play in the interactions between states, at least minimal rationality principles must be shared, as well as reference, meaning and truth. Davidson's work furthermore suggested that interpretation necessarily presupposes some trade-off between meaning and truth; in particular it presupposes that familiar semantic hypotheses will be made about the meanings of the terms used in the to-be interpreted language.

Now there are various ways in which this antecedence of the thinker's thought in the process of interpretation can be understood.   There seems to be a natural way to understand this procedure in terms of how *language use constrains* meaning and reference, rather than as a feature of a *subjective* grasp of language. Being subjective is obviously a property of a thought *as entertained by someone*. But there are objective features explaining why the thought is entertained, and that reach deeper in the causal explanation of intentional states than the brute fact that it is being thought *by me.* In particular, thought has definite informational/representational properties, that allow it to be generated, semantically evaluated, communicated, etc., given the epistemic and motivational properties of the thinker(s) involved. Finding that something makes sense or not thus does not seem to *result essentially* from a subjective impression; it may be characterized through sentence and speech analysis in syntactic, in semantic or in pragmatic terms. Such an understanding is inferential if anything is, and the inferences may be considered independently from a subjective stance. In this view, subjective appraisal of what makes sense is not a primary fact from which philosophical examination should start. It is a fact derived from properties of representational systems. Note that our reading consists in using a denial of claim 2 to weaken claim 1, just as claim 1 draws its full strength from a full adhesion to claim 2.

If the claim for the constitutive character of rationality is weakened in this way, and made independent from claim 2, then it allows to appreciate as a substantial possibility that a particular subject might fail to recognize rules of inference or misapply them in the process of interpreting other speakers. Two crucial preliminary steps in epistemology will thus consist in learning from science the mechanisms through which information is generated, collected and communicated, and from psychopathology and neurology in which particular way misrepresenting and miscommunicating can occur.

Alvin Goldman[15] exemplifies such a view, in which rational justification cannot be conducted in full independence from scientific investigation: he maintains that the aim of philosophy is evaluative (epistemology, ethics, philosophy of law illustrate most clearly a normative interest in philosophy). According to him, this aim invites rather than it precludes a rational interest in scientific (descriptive) evidence on reasoning capacities. Granted that

---

[14] Stich (1990), 38.

[15] See in particular Goldman (1986).

justification is grounded in beliefs of a certain kind, and that evaluations of beliefs derive from evaluations of belief-forming processes, an epistemologist has to assess the reliability of the processes involved in belief formation and in reasoning. As Goldman notes, "Which processes are suitable cannot be certified by logic alone, but depend on properties of our basic cognitive equipment". In this task, empirical facts about social organisation may also be relevant[16].

Let us now consider our second basic claim.

Claim 2 and the subjective essence of the mental

In *Simple Mindedness,* Hornsby complains that philosophers of mind talk about action in an impersonal language that makes it easy to forget that the agents are people:

> So we find "two beliefs produce a third" standing for a person's arriving at a conclusion. We find "an experience causes a belief" in place of the idea that a person believes something because things look to her to be a certain way. We find "pains result in avoidance reactions", substitution for a conception of a person's behavior as intelligible in the light of the fact that she suffers pain" (Hornsby, 1997, 157).

One may agree here with Hornsby: confusions between informational and subjective levels should be avoided[17]. There may be a good reason however to use a subpersonal language instead of the personal one when it comes to expressing representational properties. While the philosophers under Hornsby's critical review might perhaps express more clearly their arguments, they may still have a point using subpersonal rather than personal properties. What is needed in Hornsby's argument is a proof that *all* the facts relevant to epistemic states and agency lie within a person's subjective access and conscious control. The point again is that of the intentional closure of the subjective view. Let us develop briefly this point.

According to the subjective view, one cannot be in pain, have an emotion, form an intention, without simultaneously forming the belief that one has it[18]. One might object that pains, emotions, physical postures, intentions, often fail to be acknowledged by their bearer; the underlying states may be currently active in an individual (i.e. be contextually triggered and control her behavior in a way essentially similar to conscious pains, emotions, physical postures and intentions), while being, at least for some time, undetected by the subject. The agent herself may well acknowledge that later on, and recognize a corresponding disposition in herself in a retrospective way[19].

If an emotion, or rather a quasi-emotional state can control behavior without the agent noticing it, we have a dilemma. Should we take such a state as mental, because it controls behavior, can be acknowledged retrospectively and belongs to the kinds of states that are normally felt in a subjective way? Or should we refrain from including it in mental states, because it is currently unnoticed? These two competing features - access to consciousness or ability to control - seem to make it worthwhile to try articulating in more detail a theory of the kind of access that an organism has to her own proprioceptive and intentional states, thus

---

[16] What Goldman calls "social epistemology" is concerned with the truth-getting impact of different patterns and arrangements of social intercourse.

[17] A similar complaint is voiced by thinkers of a different inclination, like Dennett (1969).

[18] see Shoemaker, (1996).

[19] Naomi Eilan (1997) : "It is unfair to claim that "our personal level concept of remembering is wholly independent of any reference to the mechanisms that make remembering possible. It fails to do justice to the passivity involved in asking oneself questions, for the passivity here is partly a function of the fact that in doing what one is doing (asking and answering the question) there is something going on that makes this possible that one does not have access to. (..) The fact that there is this non-conscious cognitive ingredient just is what introduces the passivity."

grounding conscious or currently unconscious knowledge of one's current deeds, intentions and well-being in specific informational facts.

One might also object to Hornsby's attempt at recapturing the agent's overarching role with respect to his/her mental life that the sense of being the thinker is and that thought falls short of occupying a temporally continuous stream of consciousness; many of the subject's states are of an elusive kind, not states that a subject really controls or experiences as her own, but states that occur within her with varying results (either positively helping her solve a problem, as in directed memory and in reasoning, or as intruding her personal sphere, as in obsessive thoughts). Even in the course of controlled thinking, there is a wide spectrum of mental states that are passively entertained by a thinker, instead of being the result of controlled attention. In remembering, for example, there is clearly a contrast between the decision to try and remember, and the actual experience of retrieving from memory. Remembering is a good example of the necessity of articulating personal experience with subpersonal mechanisms in order to account for the passivity experienced by a rememberer, who depends on them without knowing which they are and how they work. Should all these psychological events in which an agent is passive towards her mental life be counted among the "external' or the internal sphere?

Some might certainly argue that as far as these mechanisms are studied scientifically, in an '"objective" perspective, they fail to coincide with what a subject experiences from the inside. In this light, therefore, there is apparently a decisive "aspectual" gap between what the subject feels as a passive experience where she allows her brain to work towards a solution, and the active operations of the brain described in theoretical psychology and in neuroscience. But such a gap consists in the unavailability of a description of the same phenomenon common to the scientist and the subject, not in a complete absence of a common *reference* between the two descriptions. What guarantees this reference is the capacity of a subject to come up with an indexical reference to all kinds of passive mental operations for which he may indeed have no description and only an indirect marker. In the case of an experience of passivity, for example, the subject may still refer to her "tip-of-the-tongue" impression, or to her "letting the name of that person surface". The subject who tries to remember a name knows that she does not know the mechanism for memory retrieval, but knows however that somehow she will manage to retrieve it; she spontaneously describes her experience as one of passivity, as one of "letting things follow their course". Another example of mental passivity is experienced when letting "sink in" an information of a radically new and potentially threatening nature: in such a case, the subject experiences that time is needed for her system to draw the needed inferences before she can face the new facts and handle the emotions involved. This sense of "extraneity" is thus quite common outside psychiatry. Any subject, "sophisticated" or not, is clearly able to refer to these episodes of her mental life through indexicals, referring possibly first to episode tokens, and later on to types of them.

If this line of reasoning is correct, expert knowledge does not need to conflict with a subject's ability to justify rationally her own plans, beliefs and actions. Even though the subject does not possess the explicit knowledge that the expert has, she is in a position to justify her own mental concept use, or the way she derives her decisions from her past experiences on the basis of her procedural knowledge of how her thought develops, through more or less controlled mental processes.

We do not want to deny that there are two faces for each mental process; on the one hand it develops as a more or less controllable procedure, identifiable indexically by a thinker; on the other hand it can be studied as series of informational processing or neuronal activation. We do not deny either that the contrast procedure/process parallels the contrast

subjective/objective. What we deny is that the contrast has a deep metaphysical significance and, in particular grounds the subjective as such. Just as one can interpret Jackson's puzzle (in Jackson, 1986) about what Mary knows as illustrating a distinction between what Mary knows propositionally and the procedural way in which Mary displays her knowledge, one can interpret the subjective-objective opposition as two sides of the same procedural coin (put to use or studied as a sequence of informational processes).

Claim 3 and the causal specificity of a common-sense view on action and subjectivity.

According to claim 3, there is a complete dissociation between the causal mechanisms that are activated when we think and act and the causal relations between our intentional states and our actions. When a person explains her actions or justifies her decisions, she need not invoke any kind of mechanism taking place inside her.

A first general remark on the relationship between commonsense psychology and the science of the mental is that the former incorporates in fact bits and pieces of "objective knowledge". It includes, beyond general rational principles and the famous "platitudes" concerning the relations between beliefs, desires and actions, empirical hypotheses about the workings of the mind and the possible causes for sub-optimal functioning. Commonsense psychology includes inter alia at least some kind of empirical explanation for *why* somebody is bright, arrogant, macho, deluded, lunatic etc. The kind of explanation used ranges from myth and tradition to vulgarized scientific results. Commonsense psychology is thus not ordinarily used as a complete body of knowledge, but rather as a field of inquiry open to extensions and enrichments from all kinds of sources. Of particular importance regarding our previous discussion of the role of deference to experts in justification, commonsense psychology includes recognition of its own incompleteness through various more or less deferential procedures that deserve careful study[20]. Even "simple-minded" people are ready to recognize that some types of behavior or some personality trait cannot be accounted for in ordinary terms. Reference is made to "conditions" or "natures" which are presented as in need of an explanation by some expert or through an ideally complete science of the mental. Nor is such a deferred qualification restricted to third-person attribution. It does involve first-person self-knowledge as well. This makes a view in which the project is to insulate commonsense psychological reasoning from empirical-scientific concerns difficult to defend in its own commonsensical terms[21].

This remark introduces a way to respond to the third claim, i.e. the claim that there is a level of causation at the intentional level (characterized in subjective terms), over and above the level of events as characterized by physics and the neurosciences. Philosophers of mind like McDowell and Hornsby consider that no mechanism story needs to be told at the personal level, and that the personal level is all there is to explore when it comes to understanding content and rational justification. A main objection is that one can acknowledge that the philosopher's task is epistemological and logical rather than purely descriptive and psychological, while also recognizing that a philosopher needs to

---

[20] See Sperber (1993), Recanati (2000).

[21] Such an incompleteness should not lure *philosophers* into accepting a double standard about the mind, with an exoteric level of commonsense   retained in philosophical explanation and an esoteric level of scientific theory that would be the privilege of a few experts. The ultimate reason why a double standard is untenable is that, even if one rejects any *obligation* of using science, the notion of a reflective equilibrium in the various fields constituting knowledge forbids *flat contradiction* with science.

characterize the epistemological and logical roles of subpersonal states and events, to the extent that the latter turn out to have a causal role in rational reasoning and self-knowledge. Moreover, as we saw in the preceding section, a person is able to refer to her own subpersonal states through indexicals. She is thus able to bridge the gap, so to speak, between her higher-order states for which linguistic descriptions are available, and the subpersonal states that are represented in her language and in her thought through indexical/deferential expressions. For example, it seems impossible to deal adequately with the justification of memory reports without taking into account scientific explanations for illusory memories, nor with the justification of self-identity without bringing to bear scientific evidence concerning both the acquisition and the pathological disruption of self representation. The idea that there is one philosophically relevant source of intelligibility, having to do with the personal level of explanation, seems thus be defeated both by the *de facto* absence of boundaries between a person's beliefs (concerning herself, other agents, as well as the physical world), and by the *de jure* necessity to strive for a coherent picture of the world, whether of the subjective or of the objective kind.

Admitting subpersonal explanation however implies a temptation for considering that the *only ontology* that needs to be taken into account is the subpersonal one. In other words, there is only one step from going subpersonal to falling into eliminativism. Dennett's early book *Content and consciousness,* for example, defends the view that two competing accounts should be offered for mental items. This would lead fragmenting mental predicates into two classes. For example, two predicates of awareness should be defined, on the following lines:

(1) A is aware$_1$ that p at time t if and only if p is the content of the input state of A's speech centre at time t.

(2) A is aware$_2$ that p at time ti if and only if p is the content of an internal event in A at time t that is effective in directing current behavior (Dennett, 1969, 118).

Awareness$_1$ would thus cover introspectible conscious states, whereas awareness$_2$ would include those brain states through which an organism represents its environment, and thus becomes able to control adequately its behavior.

Such definitions are supposed to "bridge the gap" between the personal and subpersonal levels of explanation. But do they? They may be taken rather to *expose* the gap: there is no way to identify (1) and (2) in any sensible way, because there is no person causally involved on the right side of the explanation in (2). According to Dennett, referring to personal level mental entities in a way misses its target.

> Starting from the position that thought, being what-is-reported, cannot be identified with anything in the sub-personal story, it would be poor philosophy to argue further that there must really be something, the thought, that is reported when it is true that I am reporting my thoughts (..) There is no entity in the perceiving machine, and by analogy, in the human brain, that would be well referred to by the expression 'that which is infallibly reported by the final output expression', and this is the very best of reasons for viewing this expression and its mate, 'thought', as non-referential. (Dennett, 1969, 113)

Dennett thus denies that the first personal level has to involve ontology of its own. There is only one level of causation, the subpersonal one, and the personal level, which is proper for philosophy, -- on this Dennett agrees with McDowell and Hornsby -- is also a subject matter *unsuited for causal purposes*. Some terms of the personal vocabulary may refer to actual entities at the subpersonal level; some of them however will fail to refer, for lack of a minimal matching between what is reported and what occurs subpersonally.

What is to be done with terms like "thought", or any other term (some philosophers would inscribe here: self, free will, or belief) that are found not to coincide with any item in the causal brain machinery? They are not deemed meaningless or nonsensical for that (they even

may play major roles in communicating and influencing behavior). Dennett recommends a semantical procedure, which he calls 'fusion':

> "Once we decide that a term is best viewed as non-referential, we fuse it in its proper contexts, as with 'sake' in the irreducible idiom 'for the sake of'. The contexts maintain their significance but are not subject for further analysis. (Dennett, 1969, 14)

This procedure is meant to clarify the ontology of the mental; some mental terms may be retained because of their communicative value (like "for the sake of") although no attempt is done to use them referentially. In those cases, a commonsense psychological term will be found lacking reference, and shown meaningful only as part of a larger unit of analysis. (Example: "*sake"* in " for the sake of"; or maybe "belief", "commitment", etc.).

But the tension between the two poles does not seem really eased in the fusion. Let us distinguish two types of ways in which the commonsense views of the mental fare with causal-scientific analyses: they may contradict them, or they may more or less be found to supervene on specific informational processes. Thus there will be two kinds of revisions that will have to be made by a philosopher who wants to clarify a mental ontology to make it compatible with a fully rational epistemology. Let us call these two types parapersonal vs. subpersonal revisions. *Parapersonal revisions* will *remove* items previous taken to belong to a causally efficacious personal/intentional sphere to place it into the subpersonal, with a corresponding modification of the kind of control that the subjects may have on their actions, decisions or evaluations. For example, one might have to revise attributing to oneself certain kinds of intentions, and reinterpret the corresponding pieces of behavior as environmentally controlled. *Subpersonal revisions* will reciprocally confirm the causal validity of the personal-level purported entity (property, state, event) to the extent that it does supervene on a set of well-formed, dedicated subpersonal mechanisms.

Eliminativism only seems attractive when parapersonal revisions appear to be needed on a large scale, leaving a picture of the mental that is beyond recognition for an ordinary person. This solution is however a disaster for individual rationality, for there seems to be no ground left for a subject to know whether his beliefs are correct, his desires acceptable, his plans justified, whether the way he interprets others is warranted, and even whether he is an agent at all. Another possibility is to accept piecemeal revisions of the kind indicated above.

Conclusion

As philosophers, we want to account for the way in which a subject comes to a rational evaluation of a situation. In other words, the philosopher's essential task is epistemological rather than inherently descriptive. This task requires that we describe in an as objective and complete way as possible the cognitive tools on which knowledge can be attained, as well as the ways in which we may be misled by various psychiatric or neuropsychological problems. On the other hand, our main interest is in the conscious subject, the agent and thinker for the exclusive sake of which epistemological concerns can be raised meaningfully.

This dilemma imposes some kind of complex interplay between levels, of a very different kind from the double standard that the autonomy claim would encourage. The rule should be, in my view, that subpersonal mechanisms matter only when they belong to the subject's rational sphere. Part of the philosophical task consists in showing how they do. There seems to be no way available to carry out such a task without relying on the sciences of the mental, for they allow a subject to come to grip with the basic epistemological notions in a way that coheres with science and culture at large.

References

Davidson, D., (1963), Actions, Reasons and Causes, republished in Davidson (1980), *Essays on Actions and Events,* Oxford, Oxford University Press, 3-19.

Davidson, D., (1970), Mental Events, republished in Davidson (1980), *Essays on Actions and Events,* Oxford, Oxford University Press, 207-227.

Dennett, D.C., (1969) *Content and Consciousness,* London, Routledge and Kegan Paul.

Eilan, N., (1997) Perceptual Intentionality, Attention and Consciousness, in A. O'Hear (ed.), *Current Issues in Philosophy of Mind,* Cambridge, Cambridge University Press, 182-202.

Goldman, A.I., (1986), *Epistemology and Cognition*, ,Cambridge, Harvard University Press.

Goodman, N., (1965), *Fact, Fiction, Forecast*, Indianapolis, Bobbs-Merill.

Hatfield, G., (1990), *The Natural and The Normative, Theories of Spatial perception from Kant to Helmholtz*, Cambridge, MIT Press.

Hornsby, J.,, (1993), Agency and Causal Explanation, in J. Heil & A. Mele (eds)., *Mental Causation,* Oxford, Clarendon Press, 161-188.

Hornsby, J., (1997), *Simple Mindedness,* Cambridge, Harvard University Press.

Jackson, F., (1986) What Mary did not know, *Journal of Philosophy,* 83, 291-5.

McDowell, J., (1985), Functionalism and anomalous Monism, republished in J. McDowell, (1998), *Mind, Value and Reality*, Cambridge, Harvard University Press, 325-340.

McDowell, J., (1994), *Mind and World,* Cambridge, Harvard Univesity Press.

Nagel, T., (1986), *The View from Nowhere,* Oxford, Oxford University Press.

Recanati, F., (2000), *Oratio Obliqua, Oratio recta; An Essay on Meta-representation*, Cambridge, Mass, MIT Press. Part. VI.

Rey, G., (1997), *Contemporary Philosophy of Mind*, Oxford, Blackwell.

Rey, G., (2001), A Plea for Substantive Philosophy of Mind, in C. Gillet & B. Loewer, *Physicalism and Its Discontents,* Cambridge, Cambridge University *Press.*

Rawls, J., (1971), *A Theory of Justice,* Oxford: Oxford University Press.

Shoemaker, S., (1996), *The first-person perspective and other essays* , Cambridge, Cambridge University Press.

Sperber, D., (1993), Interpreting and explaining cultural representations, in G. Palsson (ed.), *Beyond boundaries : understanding, translation and anthropological discourse*, Oxford, Berg, 162-183.

Stich, S. P., (1990), *The Fragmentation of Reason,* Cambridge, MIT Press.