Externalism, content, and causal histories

Buekens, F.A.I.

Published in:
Dialectica

Publication date:
1994

Link to publication in Tilburg University Research Portal

Citation for published version (APA):
Buekens, F. A. I. (1994). Externalism, content, and causal histories. *Dialectica*, *48*, 267-286.

Externalism, Content, and Causal Histories

Filip Buekens[*]

Summary

Externalism in philosophy of mind is usually taken to be faced with the following difficulty: from the fact that meanings are externally individuated, it follows that the subjective character of mental states and events (their accessibility for the person who "has" them) becomes problematic. On the basis of a well-founded approach to similar problems in the philosophy of action, I propose a solution based on two connected issues: (a) we should think of mental states not as beliefs, but as (defeasible) states of knowledge, and (b) thought experiments, designed to strip off the contribution of the world from the subject's contribution to the contents of his mental states, are doomed to fail. The allegedly subjective character of propositional contentful states (beliefs, desires, meanings) is that they are agent-specific states. Agent-specificity is not in contradiction with mental states or intentional actions having a circumstantial nature.

1. Narrow content versus broad content

A well-known argument for internalism points to a counterintuitive element in Hilary Putnam's famous "Meanings ain't in the head"-dictum (see Putnam: 1975). If meanings are not in the head, we cannot account for the subjective features of mental events. The underlying assumption is obvious: location of meanings plays a central role in the explanation of the well-known and often discussed asymmetry between self-knowledge (knowing what your own words mean) and knowledge of other persons' words and thoughts.

From the thesis that the meaning of indexical terms and natural kind-words is determined by external objects (for natural kinds; their essence), Putnam infers that a speaker does not necessarily know the meaning of the words he's using. (Putnam no longer accepts what he has written in his famous 1975 paper, but I'm using his arguments here merely as a starting point for a more fundamental issue.) "Narrow" content is located in the head. It is the object of a solipsistic conception of psychology. The "broad" content of a

[*] University of Tilburg, Departement Of Philosophy, PO Box 90153, 5000 Le Tilburg, The Netherlands.

word like “water” will say that it refers to H_2O. In a number of influential papers, Tyler Burge has given the argument a social twist: the content of “that”-clauses depends not only on the natural and physical environment, but on the social environment too, e.g. on the linguistic practices of the community to which the speaker belongs (see Burge: 1979).

If broad content and narrow content are indeed two different types of content, we must accept that a speaker does not necessarily know what he means by his words if the meaning of his words (the truth-conditions of sentences) are determined externally. This thesis can be deduced only if one assumes that location and knowledge of meaning are conceptually or empirically connected. If meanings are not in the head and if a speaker therefore does not always know what he means by his words, knowledge and location of meaning are indeed highly related issues.

Before I point out a *non sequitur* in Putnam’s argument, I would like to draw attention to a similar issue in the philosophy of action and its well-established solution proposed by Anscombe (1957) and further developed in Donald Davidson’s seminal paper “Agency” (in Davidson: 1980). I will then show how a similar strategy may shed new light on the above-mentioned issues.

The problem was this: what is the relation between “Jack’s moving his arm”, “Jack’s cutting with his knife” and “the killing of Jack’s wife”?[1] We have good reasons to assume that these are different descriptions of the same action. Actions are bodily movements describable in terms of their intended and unintended effects. Described as a bodily movement (“moving his arm”), we may learn how the agent has done what he did, but the description does not assume the existence of external effects. Describing the same action as “killing his opponent” assumes a causal relation between the bodily movement and the death of a particular person. We learn what he did, but the description doesn’t shed light on how he did it.

From the fact that the bodily movement is described as “killing his wife” only if a particular tragic effect takes place, it does not follow that the action is located outside the body rather than being a bodily movement. The weaker thesis that the action extends itself outside the body as a variable entity (Joel Feinberg’s famous “according principle”) is not acceptable either. What happens can be explained as follows: the bodily movement obtains the property “killing the opponent” when the external effect, caused in the right way, occurs (just as Xantippe obtains the property “being the widow of Socrates” when Socrates dies). “Killing his wife” is a relational property of the action.

[1] The question goes back to a problem raised by E. Anscombe (1957).

Given this solution, various paradoxes in the philosophy of action are easily solved. The most famous is the well-known "time of a killing"-paradox, raised by J.J. Thomson: suppose X has been shot on January 1st, but he dies the 4th. When was X killed? The (obvious) answer is that X was killed on January 1st, but that this action can by described as "killing X" only when X is dead. This solution was proposed by Davidson in his famous paper "Agency".

What follows from this approach for Putnam's dictum that meanings "ain't in the head"? Putnam's thesis seems to rest on the following assumption: if the meaning or content of certain words is partly determined by external factors, content must be located externally. But from the fact that a property of an event located *in* the head depends on what is *outside* the head, it does not follow that the mental event itself is relocated, just as the fact that intentionally killing the enemy (a property of a bodily movement, depending on the occurrence of external events) does not relocate the action (it remains a bodily movement). Similarly, from the fact that content-descriptions of events in the head presuppose that causal relations of those events with external objects and events obtain – that they have a particular causal history – a relocation of internal events does not follow. This solution, by the way, refutes Colin McGinn's claim that externalism is not compatible with an identity theory of the mind. McGinn writes:

> "It is true that identifying the mind with the brain, and mental states with the brain, conflicts with externalism, since the brain really does lie literally within the head. So it is a consequence of externalism that this simple kind of materialism is false." (McGinn: 1989, 15)

McGinn seems to endorse this view, but the fallacy in the argument can be easily demonstrated. Externalism entails that mental or intentional descriptions are relational. Relational descriptions do not relocate anything and the identity-theory of mind is not endangered. Relational descriptions connect entities with other entities, the connection being based on specific explanatory principles. When we are dealing with contentful events, those principles are the ones we use to make sense of persons and their behavior.[2]

This is not the end of the story, however. The occurrence of external intended effects is not a sufficient condition for redescribing bodily movements as intentional actions. In the case of a person intentionally killing his wife, the death of the victim is not sufficient to allow a redescription of the bodily as a killing. The death of the victim must be causally related "in the right way" to

[2] See Davidson (1984) for the well-known details.

his bodily movement.[3] The well-known problem of deviant causal chains shows that the actual causal connection between the bodily movement and the external event plays a crucial role as to whether or not we are allowed to redescribe a bodily movement as a particular intentional action. Dan Dennett offers a famous example: suppose I try to kill my adversary by shooting him. My loud shot misfires, but it causes a stampede of bisons killing the opponent. Although an action of mine caused the death of the victim, I did not intentionnally kill him (I merely caused his death).

It is by now a well-established fact that deviant causal chains may suspend the intentional description of an action. One lesson to be learned from this is the importance of the causal history of an event (its causes and its effects) when characterizing it as an intentional action. The intriguing element in deviant causal chains is that elements that do not contribute to the contents of a mental event will, if absent, suspend the intentional description of that event. I will come back to this issue.

Here is another example. When we describe someone as knowing that there's a glass of water in front of him, the cognitive state must be caused in the right way by a glass of water (not by a glass of XYZ) even if H_2O, the well-known chemical structure of water, does not figure in the causal explanation of his belief. By the same token, the belief must be caused in a non-deviant way, although we do not mention (and cannot mention) all the events (including other beliefs of the agent) that were causally responsible for that particular belief. When trying to make sense of other people's words and beliefs, we are not supposed to know everything about the world and the causal connections between events, objects and the corresponding mental states. Neither are we supposed to be able always to distinguish natural kinds from artefacts or objects that look deceivingly similar, but if the belief is caused by a hallucination caused by a real glass of water, we will disqualify this person as knowing that there's a glass of water in front of him. It is impossible to give an a priori description of what would be the "right way" in which the belief is caused and we cannot allow that all the properties of water enter in the causal explanations of his knowledge that there is water in front of him, but deviant cases or different substances will affect our description of mental states.

To sum up: what makes the sentence "X intentionally kills his wife" or "X intentionally drinks water" true (in a loose sense of "making true") is not a particular bodily movement or the occurrence of the intended effect, but a

[3] And even that is not enough: the bodily movement must be caused, in the right way, by the agent's beliefs and desires. See Davidson: "Actions, Reasons and Causes" in Davidson (1980).

complex causal configuration of events: starting with the agent's reasons and ending with that intended effect. No particular singular event or state can be pointed at as constituting a successful intentional action. Success is determined by the full causal history. Deviant causal connections or replacements of objects by things with a similar macrostructure but different chemical properties will affect the validity of the original description. I shall now investigate *how* they affect that description.

2. *Narrow content: another analogy with actions*

Fodor and Searle, among others, add to their defense of internalism the thesis that, if meanings were not in the head, we could no longer account for the association of contentful event with a conception of the world, a subjective point of view. Connecting contents with internal representations (the paradigm Fodor adheres to) makes internalism a central issue in the philosophy of mind, because it touches the very heart of mental phenomena: if content is based on a conception of the world, it can no longer be the case that ascribing content on the basis of particular causal-explanatory relations between minds and their environment (the explanation being based on principles such as rationality, coherence and truth – carefully explored in the writings of Davidson, David Lewis and Daniel Dennett, among others), reveals an essential feature of intentional phenomena. For internalism, semantic evaluation or causal relations with external events are irrelevant. This is the basis of Fodor's "methodological solipsism" (see Fodor: 1980), Stephen Stich's "principle of psychological autonomy" (see Stich: 1983) or, to a certain extent, Searle's "rediscovery of the mind".[4] I shall try to show that this is not the kind of internalism we need defend if we assume, *contra* Putnam, that meanings are located in the head. On the contrary: a causal-historical view on mental events (that locates them in the head), denies that there are, in any philosophically interesting sense, internal or external entities that confer content upon mental states. This form of internalism in philosophy of mind is as innocent as the internalism accepted in philosophy of action: Intentional actions are bodily movements, but there is no intrinsic property to be associated with the bodily movement that tells us what the agent intentionally did or that accounts for acting intentionally. In order to qualify his bodily movement as an intentional action, we must consider the causal history (the causes and effects) of that movement.

[4] I am aware of the fact that these authors have very different, more sophisticated motivations for their projects.

Before I develop these points further, let me explore another analogy with actions. The relations between bodily movements, their external effects and intentional actions help explain the relation between success and failures, between trying to act and intentionally acting. Actions succeed only if the external, intended effect occurs and is caused "in the right way". We cannot reduce intentional action to intentionally trying to act (where the trying only involves a bodily movement).[5] Contrary to many theorists, I do not think the notion of "trying to act" plays a central role in the theory of action. That actions succeed most of the time and that bodily movements have causal connections with external events and states are constitutive assumptions if we want to make sense of actions. "Trying to A" (where "A" is the description under which the action, if successful, is intentional) is from this point of view not an entity that precedes and/or causes every intentional action, but an intentional action under a loaded description, either because we're not yet sure whether the external, intended effect will take place or because we know that the action failed ("he tried to hit the bull's eye, but missed it"). If this is true, we can no longer claim that tryings are entities common to successful and unsuccessful actions. If a person acts successfully, we normally do not say he tried to act – we say he acted successfully.

From the empirical possibility that a particular action can fail, it does not follow that all our actions may fail. The norm for action is: successful action, just as the norm for belief is: true belief (see Davidson: 1984). If an action fails, we can say we tried to act. "Trying to A" does not refer to something common to successful and unsuccessful actions; it says that, in this particular case, the relevant effect didn't occur (or, in a slightly different context, that we are not sure the relevant effect will occur). To conceive of tryings as entities which cannot fail is based on the false assumption that what is immune to failure must play a central role in an account of acting intentionally. Our understanding of actions depends on the assumption that most of our actions succeed. Only under that assumption can we make sense of someone's actions (including those that failed). This general statement doesn't tell us (and shouldn't tell us) which particular actions are successful; nor does, in a different context, the general statement "most of our beliefs are true" tell us which beliefs are actually true. Describing a particular action as "trying to A" has the same status as describing ourselves with the words "I believe that p" when we

[5] Some authors, when confronted with this claim, immediately realize that not even the bodily movement has to be within one's reach. For them the trying is thus an event that precedes and causes the bodily movement; see Hornsby (1980) for this theory. I develop an alternative that is in line with this paper in Buekens (1992).

are not sure as to what is the case: it tells us that we are not certain that the belief is actually true.

One might be tempted to assume that tryings or attempts play a central role in a theory of action on the basis of the following argument: external effects fall outside the "scope" of the agent's control; although it is fully within the reach of a competent agent to move his body in a certain way, it is not (or no longer) up to him whether the external, intended event occurs. From the fact that in particular cases actions may fail because the external events do not occur, it is inferred that the ontology of intentional actions must be restricted to reasons and bodily movements (or, in the light of the fact that bodily movement may fail, to something "inside" the body: the attempt to act).

But that is, once again, a *non sequitur*: just as it doesn't follow from the fact that beliefs may be false that external events are to be excluded from a theory of cognitive states, it doesn't follow from the fact that external effects of bodily movements (or bodily movements themselves) may not occur (actions may fail), that they are to be excluded from a theory of action. As Davidson has argued, that most of our beliefs are true is a well-motivated normative statement; it cannot be falsified or rejected simply because false beliefs occur. The same is true for its equivalent in action theory: that actions are mostly successful is a normative claim; it is not refuted by the fact that particular actions can and do fail.

What lies behind these attempts to formulate a coherent form of internalism is the tendency to see externalism as a theory that jeopardizes the autonomy of the agent. The tendency to look for something common to successful and unsuccessful actions, and to think of this element (the attempt to act, the trying) as something that cannot fail or something one cannot be wrong about rests on the mistaken idea that all the elements on which intentionally acting depends must fall under the agent's cognitive control. Acting successfully is a circumstantial matter. It depends on causal chains with external events. A theory of action must study the agent *in situ*. An *in vitro* theory of intentional action does not contain the necessary elements to understand the idea of acting intentionally. If action is circumstantial, then cognitive states must have that property too, or so I shall claim.

3. Believing and knowing, trying and intentionally acting

I shall now further develop the analogy between acting and trying to act, and knowing that p and believing that p. The claim is that cognitive states, like intentional actions, depend for their existence on causal connections with external events. Once this is established, I will show that the distinction between narrow content and broad content is baseless. Thought experiments designed to support this distinction merely reveal that, if content is based on causal connections with external objects and events, attribution of content depends on non-semantically relevant elements. This is the counterpart in philosophy of mind of giving up the idea of autonomy in philosophy of action.

First, the analogy. The idea that externalism is a contentious conception of mind seems to depend on the following considerations: beliefs have content and can be evaluated semantically (they are true or false). But the fact that they can be false or can deal with inexistent entities (unicorns, square circles and other exotic entities) without losing content seems to imply that content cannot simply depend on relations with external events. True and false beliefs must have something in common: an abstract proposition, the intentional object they are "directed at" or its more popular variant: a representation couched in a language of thought. Content cannot be determined by what is the case is a belief is true, otherwise false beliefs would lack content (which they do not).

We are misled here by two different uses of the locution "X believes that p". There is, first, the use of the two-places predicate "to believe" in the specification of the content of a cognitive state. It presupposes the existence of a mental state and a sentence connected to it by means of the predicate "believes that". (In this, it functions exactly the way the predicate "weighs" connects objects with numbers.) Secondly, there is the use of the predicate "believes that" as an epistemic qualifier. When we say that X believes that p, giving the story an epistemic twist, we say that he has a certain conception or representation before the mind, but we emphatically do not want to say that what he believes is true, that what he represents with what is the case, etc. We are not committing ourselves to the truth of his belief. The content of his belief is the way he represents the world (and not how the world presents itself to him), what is true from his subjective point of view. This leads to the erroneous idea that mind can be studied *in vitro*, independently from the causal history of the states and events that constitute mindful beings: to specify the content of one's belief, we specify not what they represent, but the representation itself. Not that the content-specifying role of "X believes that p" is, by itself, not sufficient to call into life representations. The additional element needed to introduce these entities is based on an epistemic gloss: what we believe rests upon a (true or false) conception of the world.

The epistemic gloss inspires the so-called extraction problem, an issue well-known to proponents of internalism.[6] From the truth-conditions of the

[6] See for instance M. Devitt (1990).

sentences we use to describe the content of one's beliefs, we must extract how the agent represents reality. Michael Devitt and Brian Loar are among the authors who have written extensively about this problem. A scientific solipsistic psychology will study the content of beliefs, having stripped it from its truth conditions or the external states and events they are about, just as a scientific theory of action should abstract away from the external contingencies that accompany intentional actions (a proposal made in Stich: 1983).

The pressure to introduce something that is common to true and false beliefs comes from a misleading but tempting conception of the mental: something that must be in the mind even if the outside world had a totally different outlook. The counterpart in the philosophy of mind of the philosopher of action's autonomous agent – the idea that acting simply means: moving your body in the right way, or trying to act – is the idea of having a representation before you, independent of the way the world looks. The difficulties with this view are mirrored in the internalism that introduces representations. It is not because what is described on the basis of their causal history and by invoking specific explanatory principles are internal events, events *in* the head, that we can assume that the described states can be studied independently from that causal history and the principles that make that history available to us.

How can we avoid the epistemic gloss associated with "believing that p"? I propose a move that may have an air of paradox (at first sight): a switch from belief to knowledge.[7] The move is paradoxical because we assume that "knowing that p" is "stronger" than (merely) believing that p. If X knows that p, he not only believes that p, but he also has the right reasons, he is justified in believing that p. Exploiting a suggestion made by Bernard Williams, I think we can reject this connotation. "I am represented as checking on someone's credentials for something about which I know already. That of course encourages the idea that knowledge is belief plus reasons and so forth. But this is far from our standard situation with regard to knowledge; our standard situation with regard to knowledge (in relation to other persons) is rather that of trying to find somebody who knows what we don't know; that is, to find someone who is a source of reliable information about something." (Williams: 1973, 146). The assumption is that someone is a reliable source of information about a particular fragment or aspect of the world only if there is particular causal link between this person and the fragment or aspect of the world. A similar idea is defended by McDowell (1980): communication is more than

[7] A similar move (for slightly different purposes) was proposed by S. Guttenplan at the Neuchâtel conference (see his paper in this issue). I thank him for additional remarks at the conference.

simply making available your beliefs to others or manipulating other person's beliefs about your beliefs. It is the instilling of knowledge. In communication, knowledge spreads like a contagious disease (a phrase we owe to Gareth Evans).

The kind of externalism I try to defend in this paper is not, of course, that we relocate representations and think of them as external facts. Thinking of cognitive states as knowledge states (and not merely as belief states) reminds us of the platitude that, when interpreting persons, the explanatory principles governing that enterprise force us to consider such states as true and coherent. Describing them as states of knowledge (and not mere belief-states) presupposes that we consider the person they are ascribed to as someone who carries information about the world. This is not to say that he cannot have *mere* beliefs – where describing a cognitive state as a belief state is a loaded description of that state, either because we're not sure as to the truth of that state, or because we know that what we ascribe is false (cf. supra). Many discussions of the principle of charity deal with the kind of problems we encounter when confronted with someone holding sentences to be true that are obviously false: what he thinks of as a state of knowledge is a mere belief. As interpreters, we must assume that what we describe are states of knowledge, notwithstanding the fact that particular mental tokens may be mere beliefs.[8]

Belief and knowledge are related to each other the way tryings are related to successful intentional actions. If something goes wrong in the external causal chain (or if we expect something to go wrong), we can produce a loaded description of that state: we say it's mere trying, a mere attempt, or a mere belief (not a state of knowledge). The standard case is the state that is connected in the right way with external events and objects: an intentional action or a state of knowledge. A belief is thus not arrived at by *subtracting* something from the causal chain that results in a state of knowledge, or by observing the internal part of that chain. It is arrived at by observing an alternative causal chain.

The kind of externalism that conceives cognitive states as states of knowledge has no problem in admitting that they have causal connections with external events. What is crucial here is that we are no longer tempted to introduce representations or points of view, the kind of things that have content or constitute content independently of their connections with the outside world. Cognitive states, thought of as knowledge-states, exist in the light of complex causal connections with external events. It is the causal chain leading to that particular state that is responsible for our describing that state as, say, know-

[8] See also S. Guttenplan's paper in this volume for more arguments that support this view.

ing that there's water in front of you. Deviant causal chains or misleading circumstances (H_2O/XYZ-cases) may prevent us from ascribing knowledge to a person. Instead, as we shall see in a moment, they may force us to describe them as something less: mere belief states.

Let's take stock. We have reached two conclusions: (a) elements in the causal history of a mental event that do not contribute to its contents may nevertheless, when absent, affect the validity of the content-ascription; (b) the description of mental states requires that we account for them as knowledge states. We give up this default assumption and characterize them as mere beliefs when we are in doubt about their truth-value, or when we know that elements in the causal history of that state prevent us from putting the agent in the position of someone who knows what's going on.

If the relevance of causal chains must be acknowledged for when describing cognitive states or intentional actions (thereby accepting that externalism is correct) we are in for a much stronger and probably more controversial claim. I shall dub this position *actualistic externalism.* Actualistic externalism claims that thought experiments designed to draw a distinction between narrow and broad content, between the contribution of the subject and the contribution of the world to the contents of a mental state, have no real significance.

4. Actualistic externalism

In the first part of this paper, I stated that elements which are not relevant in identifying the content of an event but nevertheless determine its identity, may affect the truth of the description under which it is intentional. I will now elaborate that thesis and then connect it with issues from the foregoing section.

First, we have the claim that any change in the causal history of a mental event (by replacing it by a physical duplicate, by changing its environment, by introducing deviations in the causal chain) changes the identity of that mental token. This is trivial in the light of Leibniz' principle that the identity of an object or event is determined by all its properties (relational, intrinsic or whatever) and the relevance of the history of an event for the properties it has or the descriptions that are true of it (I hope nothing in this discussion depends on the question whether we accept either properties or predicates that are true of objects).[9] Properties can be relational. Any change in the history of an event may add or withdraw relational properties and thus change its identity.

[9] Pascal Engel has persuaded me that difficulties are lurking here.

This claim is not meant to imply or to suggest that there are no other features that determine the identity of an event. It merely states that, among the properties that determine its identity, some are historical, i.e. determined by its causal history. It is obvious that not everything that determines the identity of a mental event contributes to its content or figures in the content of that event.

The second claim is more contentious: rational explanations are token-bounded. When we try to make sense of other persons, and the reasons on which they act, we redescribe token-events on the basis of well-known explanatory principles such as overall coherence and truth. Agents must be rational, most of their beliefs must be true and their desires must be reasonable in the light of those beliefs. A rational explanation is the explanation of a token-action, in the light of its overall coherence with other actions (verbal or non-verbal). Token-boundedness of rationalizing explanations means that such explanations do not start with a full-blown theory which is then applied to particular cases, but rather with an evolving theory of a person's mind, geared to particular actions or speech acts.[10] Token-boundedness entails that the intentional description a rational explanation has produced for a particular event *e* cannot be used to describe an event *e'* that is in any physical (but not historical) respect the same as the original event (this is one aspect of Davidson's theory of the supervenience of the mental on the physical). Alternative events require new (causal) explanations. Token-boundedness of rational explanations is a consequence of anti-reductionism based on Davidson's arguments in "Mental Events" and subsequent essays.[11]

From the conjunction of these claims it follows that events which have different causal histories (and therefore have different identities) be submitted to different interpretations, *even if the change in identity is due to a difference in the causal history that is not relevant within the interpretative, rational scheme we adopt in order to explain the original event under its mental description.* This is paradoxical, because we readily assume that semantic distinctions or explanations of content ought to be based solely on semantically relevant facts, facts that figure in the scheme of rational explanation we adopt when making sense of persons. It is an almost generally accepted view in the philosophy of mind and action that non-mental facts (i.e. physical facts) cannot affect the correctness of mental descriptions. Semantic distinctions must be

[10] See Davidson (1987) for this conception of such a theory.
[11] See Davidson (1980) for the details.

based on evidence visible only from within the intentional stance.[12] What is not relevant from that point of view cannot affect semantic descriptions.

The relevance of non-semantic elements is shown for intentional actions and for mental states. First, the problem of deviant causal chains. Think of Dan Dennett's assassin who missed his victim but whose loud shot caused a stampede of bisons killing his adversary. The explanation of the killing of the victim in normal circumstances would never have mentioned all the elements that could have gone wrong in the causal chain starting with the bodily movement (pulling the trigger) and ending with the death of the victim. The explanation works because it points to what is relevant from the point of view of a rational explanation of that action. However, what was not relevant in the causal explanation of the successful action will, if absent or responsible for an alternative causal chain, become relevant and prevent us from applying the original description in the deviant case.

The positive counterpart of this thesis is that no rational agent will be able to fix all the circumstances relevant to the success of his action. To illustrate this hypothesis, John Perry offers the following example:

> "Consider the force of gravity, if I am in space or on the moon . . . the movement [I normally] envisage . . . will not lead to getting a drink. The water would fly out of the glass all over my face – or perhaps I would not even grab the glass but instead propel myself backwards. If all possible failures are to be accounted for by false beliefs, the corresponding true beliefs must be present when we succeed. So, when I reach for the glass, I must believe that the forces of gravity are just what they need to be, for things to work out right. But it hardly seems probable that everyone, even those with no knowledge of gravity, believes, when they reach for a glass of water, that the gravitational forces are what they are; such an attribution would drain the word "belief" of most of its content (. . .). A more efficient way of Mother Nature to proceed is to fit our psychology to the constant factors in our environment and give us a capacity of belief for dealing with the rest." (Perry: 1986, 131)

"Dealing with the rest" is, in this context, the capacity for dealing with mental states caused in deviant ways or mental states in non-benevolent circumstances. Twin Earth experiments offer an interesting (and well-known) class of non-benevolent contexts. Suppose I am, unbeknown to myself, transported to Twin Earth. In front of a glass of twater (a liquid closely resembling water, but with a different chemical structure – XYZ), I may believe there is a glass of water in front of me, but I cannot say I know there is water in front of me. The

[12] The same is true for moral facts: we do not assume that non-moral consideration can affect the moral. But see Burms & Vergauwen (1991) for an alternative view on this issue.

non-semantically relevant causal history of my cognitive state affects its epistemic status. By the same token, we are not able to fix all the circumstances relevant to our having knowledge about the external world. From this, it doesn't follow that we cannot have knowledge of our normal surroundings. A combined example is this: suppose, being on Twin Earth, I cannot say I intentionally drink a glass of water, because there's no water in the glass I drink. Although I do not have to know all the properties of water to perform the intentional action of drinking a glass of water, replacing it by a glass of XYZ suspends the description "intentionally drinking a glass of water". I have merely attempted to drink a glass of water. (Nothing says that this description must be available to me when I perform this act.)

What lessons should be drawn from these cases? There are obvious counter-moves designed to avoid the seemingly disastrous conclusion that elements that are non-semantically relevant may suspend semantic descriptions. A well-known move consists of introducing a distinction between broad content and narrow content. The move is simple but radical: deny that the elements that affect the truth of the mental description are not semantically relevant. Simply claim that they too are part of the semantic realm: they determine the broad content of a belief, or the broad description of an intentional action. Broad content thus takes into account features that determine the truth-value of a belief (so the story goes). Although these allegedly semantically relevant features must not be known to the speaker (to be explained by their location: they are located outside the head), they are part of the broad content of the belief. (The equivalent in action theory is Feinberg's accordion theory: what I do is not only externally determined – the external elements are part of what I do.) The move is based on the following argument:

(a) The identity of an event is partly based on its causal history.

(b) Mental events with different causal histories have different identities and therefore require different causal explanations of their mental content.

(c) Because nothing outside the semantic realm can affect semantic descriptions, we must assume that elements affecting the truth of the description are indeed semantically relevant.

Therefore,

(d) systematic changes in the identity of a mental event by manipulating its causal history correlate with systematic changes in the content of that event.

And from (d) it follows straightforwardly that

(e) thought experiments with possible worlds reveal the distinction between the subject's contribution to his mental states and the world's contribution to their content.

But (d) does not follow, because (c) is simply false. The only sensible conclusion to be drawn from (b) is that alternative causal histories require alternative explanations. We make sense of an unsuccessful action not by inspecting what is common to a trying and a successful action, but by observing the alternative causal chain. Qualifying a bodily movement as an unsuccessful action is based on an inspection of the causal history of the unsuccessful action. By the same token, qualification of a mental state as a false belief is based on inspection of its particular history.

Thesis (c) leads to the erroneous idea that we could, with the help of thought experiments that change the causal history of an event (in most cases: the external history of the event), make a distinction between the contribution of the agent and the contribution of the world to the contents of his mental states. But (c) is not true. Alternative causal histories show that we disqualify a mental state as a state of knowledge or as an intentional action. We do not conclude from these cases that there is something common to the unsuccessful and the successful action, or arrive at what must be added to an internal representation to obtain its truth-condition. (Deviancy can occur between the reasons and the action.[13]) From alternative causal chains, nothing is learned as to what an agent cannot fail to do, or what an agent cannot fail to know. Stipulating alternative causal chains does not tell us (and cannot tell us) the limits of the autonomy of the agent or which elements of content he has privileged access to.

Changing the causal history of a cognitive state comes down to creating a new token which requires an alternative interpretation. No common (internal) element is identified, and neither can we detect the systematic contribution of the world to the content of the token unless it is explicitly given when stipulating the alternative causal history, but that move trivializes the experiment. There is no specific or systematic item we can eliminate from the causal history of a state of knowledge to arrive at a mere belief. The alternative state

[13] See Davidson, "Agency", in Davidson (1980).

(a mere belief state, or an unsuccessful action) is individuated by inspecting the alternative causal chain. The relation between the history of a person we use to make sense of him and a particular deviant chain is that of a normal background on which deviant cases (actions that turn out to be mere tryings, cognitive states which are mere beliefs) are projected. That normal background includes elements which are not relevant when specifying the content of mental states, but which nevertheless determine their identity.

As alternative causal chains suspend mental descriptions, we are tempted to think that by stipulating external alternatives, we detect the contribution of the world to the content of our beliefs. And the world outside must contribute to the content of mental states on the basis of the principle that nothing can affect the correctness of descriptions of content unless it enters into the content of mental states. But that principle is simply not true: non-semantic facts, or non-contentful events do affect the correctness of semantic descriptions. When confronted with a glass of water, we normally do not mention that it is H_2O the speaker knows is in front of him; but when water is replaced by its well-known Twin Earth variant or, more mundane, by a deceivingly similar liquid, we do not say that the agent knows there's water in his glass. He merely believes there's water in his glass. The Twin Earth variant forces us to describe the agent as merely believing there's water in front of him, but it does not follow that the ("broad") content of his mental state must mention that there's XYZ out there. It simply affects the epistemic status of his cognitive state. Nothing in this story amounts to a difference in content between a state of knowledge and the content of a corresponding ("mere") belief state.

The same is true for actions. When a deviant causal chain occurs, the deviant element allows us to describe what the agent did as an attempt to A, not as intentionally A-ing. The alternative chain doesn't affect the description of the attempt. We describe the attempt using a sentence that would be true were the action successful. That description is available insofar as we are able to project the unsuccessful action or the mere belief on a background of successful actions and states of knowledge. That background is offered by the history of the person. The difference between the world in which an agent successfully acts and a world in which he fails doesn't tell us what was common to both cases and what is (therefore) supposed to play a central role in a theory that accounts for both successful and unsuccessful actions. Too many things can go wrong; nothing remains *a priori* constant among the possible variants we can imagine.

I conclude that differences in external circumstances are relevant for the ascription of content, not because they reveal a systematic semantic contribution of those circumstances to the content of a mental state (its "broad content"), but because any break with the normal background reveals the dependence of the validity of semantic descriptions on *prima facie* non-semantically relevant features making up the normal circumstances in which we acquire knowledge or act successfully.

What, then, is the relation between the intentional action and a mere attempt, or between a state of knowledge and a mere belief? Representationalists see the relation as follows: to arrive at the contents of a belief, we must substract the contribution of the world, the external circumstances, from the corresponding knowledge token (that would be a consequence of Devitt's extraction-problem – cf. supra). The contribution of the world to the meaning of "water" is that it is H_2O, and therefore one cannot be in a state of knowledge about water unless one knows *it is* H_2O or adds *to the belief that it is* a belief about H_2O. Similarly, what must be substracted from intentionally drinking water to arrive at an attempt is the contribution the world makes to that action. The description of the attempt cannot refer to water but must be something like an attempt to drink a tasteless, colorless liquid (if that is what the extraction would amount to).

I contend that it is more apt to talk of an addition problem than to talk of a subtraction problem. What must be added to a knowledge state to obtain a (mere) belief state are deviant circumstances that prevent one from being in a state of knowledge. What must be added to an intentional action (drinking water) to obtain a mere attempt to drink water are the deviant circumstances that prevent one from being successful. More important, however, is the fact that we are not able to demarcate a fragment in the causal chain that is common to both cases.

5. Externalism and the subjective character of the mental

Thought experiments which vary external circumstances play a double role. There is, first of all, their demarcation-function: they are supposed to make a distinction between truth-conditions and verification-conditions, what remains "constant" from the point of view of the agent and what makes such states true. Blackburn (1984) aptly dubs this the "spinning the possible worlds-strategy". But if my thesis is correct, we can no longer discriminate, by spinning the possible worlds, the contribution of the agent from the contribution of the world. Blackburn associates this strategy with the subjective character of mental states as follows: the purpose of spinning the possible worlds is:

> "(to) keep things as much as possible the same from the subject's point of view, while imagining different external causes of his being in the state he is in." (Blackburn: 1984, 312)

If my version of externalism is correct, Blackburn's strategy is doomed to fail. Variations in the external (or internal) causal history of a mental token affect its identity and therefore suspend the correctness of the mental description it verifies. As we have seen, it doesn't follow from this claim that the variation introduced in the thought experiment enters into the description of the content of the original or new mental token. Events with different causal histories, however similar they may be, require different causal explanations.

The circumstances we live in – those that make it possible for us to act intentionally or to be in the position of someone who knows what's going on in the world – matter a lot to us. It would be futile to say that external states, events and circumstances that figure in one's personal history and thus make up what one believes and says, do not matter, even from a so-called "subjective" point of view. They determine who we are and what we do. Who would deny that such features are not relevant from a subjective point of view? It is hard to reject the idea that whether or not one successfully kills someone (and thus becomes a murderer) is not important for him or her. The same is true for the distinction between knowledge and belief: when I come to believe something in a deviant way, or because I was transported to a different world, or, more down to earth, because the world has drastically changed, I can no longer claim that I carry information about the world. Being in the position of someone who knows his way around in the world is an all-important matter for the picture we have of ourselves. A full conception of myself must be based on more than what I do with my body or what arrives at my retina.

"But this is not an account of what I have private access to", some may object to these remarks. Indeed it is not. If the causal history is relevant to determine the content of what I do, know, try or believe, it no longer makes sense to say there is an inner realm that accounts for the asymmetry between self-knowledge and knowledge about other persons. How can we then explain the asymmetry?

Davidson has described the asymmetry as an authority problem. We accept self-descriptions by other persons as true because they play a central role in building a theory for their language and their mind. Accepting them as true plays a constitutive role when interpreting (making sense) of other persons. The authority I have is due to others deferring to me about my self-ascribed thoughts. But how convincing is this? "Suppose no one deferred to me about my self-ascribed thoughts. Would my confidence that I know what these thoughts are diminish? Do I not know my thoughts whether or not others acknowledge that I do?" (LePore: 1989, 210).

The position I have outlined in this paper contains an important element that may help explain at least one aspect of the asymmetry. Part of the argument against the "spinning the possible worlds-strategy" was that there is no point in changing the external circumstances so as to detect what remains "constant" from the subject's point of view. This seems to imply that the alleged asymmetry between the first person and the third person is a mere illusion; content is determined by what is outside the head. What I contend now is that the basis for the asymmetry between self-knowledge and knowledge of the mental life of other persons is actually rooted in the fact that it is the particular causal history of a person that reveals what he thinks and what his words mean. There is and will always be an other minds-problem, in this sense: however well I understand a person (having understood him on the basis of his history), I could never apply that explanation to a different, third person, and claim to understand that person on the basis of what I already know about the second person (or, by the same token, what I already know about myself). If interpretation is token-bounded, understanding is based on interpretation, not on the application of a ready-made theory to particular cases.

If this is true, we have an obvious reason why there is a genuine asymmetry between you and me. Every person has the privilege that he is the only one who, by interacting with the world as best as he can, makes evidence available to his interpreter. "Making evidence available" doesn't mean that he has private access to an inner realm and then, in a second move, goes public through interaction with the world; it means that he is the only one who can, by interacting with his environment, enable us to arrive at a causal explanation of what he believes. He creates his own, personal history – the history that makes it possible to determine what is, for him, a normal context, what are, for him, normal circumstances and how they shed light on deviating cases. This is what determines his authority.

What we know about our own mind is not available as a kind of theory we can apply to other persons (and neither can we apply the theory that was geared to a given person to someone else), because that would violate the token-boundedness of rational interpretations. We must, of course, *use* our own beliefs, desires and language when understanding other persons, but that is a substantially weaker (but definitely more correct) thesis than the more controversial one that we can apply a theory that makes sense of ourselves to other persons.

References

Blackburn S. (1984), *Spreading the Word. Groundings in the Philosophy of Language*, Oxford.

Buekens F. (1992), «Essayer, réussir et échouer dans une action», *Revue de théologie et de philosophie* 124, 231–248.

Burge T. (1979), «Individualism and the Mental», *Midwest Studies in Philosophy* 4, 73–121.

Burms A. & R. Vergauwen (1991), «Natural Kinds and Moral Distinctions», *Philosophia* 21, 101–105.

Dancy J. (1985), *Introduction to Contemporary Epistemology,* Oxford.

Davidson D. (1980), «Agency», in Davidson: *Essays on Actions and Events,* Oxford, 1980, 43–62.

Davidson D. (1984), *Inquiries into Truth and Interpretation,* Oxford.

Davidson D. (1987), «Knowing One's Own Mind», *Proceedings of the American Philosophical Association* 60, 441–458.

Devitt M. (1990), «A Narrow Representational Theory of the Mind» in W. Lycan (ed.) (1990), 442–468.

Fodor J. (1980), «Methodological Solipsism Considered as a Research Strategy in Cognitive Science», *Behavior and Brain Sciences* 3, 63–73.

Hornsby J. (1980), *Actions,* London.

LePore E. (1989), «Subjectivity and Environmentalism», *Inquiry* 33, 197–214.

Lycan W. (ed.) (1990), *Mind and Cognition. A Reader,* Oxford.

McDowell J. (1980), «Meaning, Communication and Knowledge», in Z. Van Straaten (ed.): *Philosophical Subjects,* Oxford, 1980, 117–139.

McGinn C. (1989), *Mental Content,* Oxford, 1989.

Perry J. (1986), «Circumstantial Attitudes and Benevolent Cognition» in J. Butterfield (ed.): *Language, Mind and Logic,* Cambridge, 129–143.

Putnam H. (1975), «The Meaning of "Meaning"» in Putnam: *Mind, Language and Reality. Philosophical Papers,* vol. 2, Cambridge, 215–271.

Stich S. (1983), *From Folk Psychology to Cognitive Science: the Case Against Belief,* Cambridge.

Williams B. (1973), «Deciding to Believe» in Williams: *Problems of the Self,* Cambridge, 1973, 136–151.