Agential Systems and Basic Causal Deviance

According to the Causal Theory of Action (CTA) an intentional action is an event caused by a mental state that rationalizes its execution. The type of event that corresponds to an intentional action is usually a bodily movement, while the type of mental state that corresponds to the causal antecedent of such event is usually an intention. CTA has traditionally been saddled with problems emerging from so-called "deviant causal chains", namely, chains of events that satisfy CTA's conditions for the production of an intentional action but whose product is intuitively not an action. A plausible strategy for defending this theory against the possibility of deviant causal chains is grounded on the proposal that the bodily movement corresponding to an intentional action must be sensitive to the content of the mental state that causes it. Sensitivity here is understood as the specific responsiveness that a bodily movement can have to the particular content of a motivating mental state. However, this strategy faces a serious challenge from cases where bodily movements are produced by means of causal chains of events that involve intermediate intentional actions performed by a second agent. These cases are quite challenging because they allegedly generate deviant causal chains that nonetheless satisfy the sensitivity condition.

Consider Christopher Peacocke's [1979: 86] fictional scenario where a neurophysiologist reads a behavioural intention directly from a patient's brain and then intentionally stimulates the patient's efferent nerves. The stimulation then causes a bodily movement that matches the content of the patient's behavioural intention. Peacocke believes that this case is a counterexample to CTA's effort to deal with deviant causal chains by appealing to sensitivity. For although the patient's bodily movement is transitively caused by an intention that rationalizes its execution and is sensitive to the specific content of this intention, it is not one of the patient's intentional actions. According to Peacocke, the reason that the patient's bodily movement fails to count as her intentional action is that its production contravenes a fundamental assumption concerning agency, namely, that the agent must be the originator of her own actions.[1]

If Peacocke is right in thinking that scenarios involving what we may call "prosthetic agents" generate causal deviance, we seem to have a serious challenge to CTA's effort to avoid causal deviance by appealing to sensitivity. But, oddly enough, if one accepts that the neurophysiologist case exemplifies a deviant causal scenario, then one also needs to explain in what sense this particular case is different from other scenarios involving prosthetic agents that share with this deviant case all their salient features without themselves being obviously deviant.

To illustrate this puzzle, consider a similar fictional scenario where an assistant's job is simply to hold the wires that connect some efferent nerves of a patient.[2] By performing this intentional action the assistant permits the patient's behavioural intentions to be causally connected to the patient's intended bodily movements. This is a case similar to the neurophysiologist's because the causal chain that leads to the bodily movements of the patient runs through the assistant's intentional action of holding the wires. And yet, in contrast to the neurophysiologist's case, it is hard to say that the ensuing bodily movements are not intentional

[1] Here is Myles Brand articulating a very similar assumption: 'A person must perform his own action: no one can perform someone else's action. A person can be guided, cajoled, commanded, coerced, even hypnotized, into acting; but nevertheless, if he acts, it is *his* action.' [1984: 22]

[2] John Bishop presents this example [1989: 159].

actions performed by the patient. Indeed, the causal role of the assistant appears to be the same as the causal role of the patient's own efferent nerves, namely, being a causal bridge connecting the patient's intentions to the patient's bodily movements.

Peacocke's reply to the challenge arising from deviant cases involving prosthetic agents consists in requiring that intentional behaviour explicitly exclude the possibility of two agents interacting in this way. He believes that by stipulating that a causal chain leading to an intentional behaviour 'should not run through the intentions of another person' [1979: 88], one can safely rely on the sensitivity condition to take care of deviance. And, of course, if by definition no cases involving two agents interacting in this way are possible, then the puzzling resemblance between the cases of the neurophysiologist and the assistant is not an issue.

But this is hardly a satisfactory reply if one is a supporter of CTA. Not only is it an arbitrary exclusion of cases that are clearly conceivable, but this reply is also in tension with CTA's general approach to intentional action, in particular, with CTA's acceptance of transitive causal relations involving actions as much as wires and nerves. [3] One would like to know exactly what is wrong or unacceptable about a causal chain of events that goes through someone else's intentional action, produces an event that satisfies the conditions to count as an intentional action, and yet fails to be an intentional action. [4]

So, unless one wants to stipulate with Peacocke that any causal chain that goes through another agent's behavioural intention is by definition unacceptable, providing an answer to the challenge to CTA that arises from the possibility of cases involving prosthetic agents turns out to be a very important task for the viability of this theory. Moreover, given the crucial role of causal deviance in establishing the causal conditions that identify an intentional action this otherwise somewhat technical issue turns out to be critical for CTA. This is exactly what has motivated defenders of CTA to try to face the challenge coming from cases involving prosthetic agents and reap the benefits of a successful answer to it. The most important of such benefits is the alleged possibility of offering the set of necessary and sufficient conditions that identify an intentional action. Here is where John Bishop's proposal to deal with this challenge becomes quite relevant, for not only it is the most suggestive effort to explain away this type of deviance but it is presented by him as the "Final Breakthrough" that would once and for all provide us with the necessary and sufficient conditions to identify an intentional action. In the rest of this paper I examine Bishop's proposal and argue that it fails in attaining its intended goal.

Bishop's "Final Breakthrough"

Bishop's version of CTA belongs to a family of causal theories that consider agents as *functional systems* capable of producing intentional behaviour by entering into internal states that besides

[3] The most common cases where an agent transitively causes another agent's intentional action are so-called "interpersonal interactions" [Hart and Honoré: 1959]. They involve a first agent causing a second agent's intentional action and hence producing a causal chain of events that, as in the present cases, also goes through an agent's mental states. However, the main difference between the present cases involving prosthetic agents and interpersonal interaction is that in cases involving prosthetic agents the relevant chain produces a basic intentional action with the help of someone else's intentional action, while in interpersonal interactions a causal chain starts with an agent's basic intentional action and ends with another agent's basic intentional action. Nonetheless, these two cases exhibit the versatility of CTA with respect to its use of causal transitivity.

[4] At this juncture one may be willing to bite the bullet and propose that cases like the neurophysiologist's are not deviant. That is, propose that the causal chain satisfies in the relevant way the content of the motivating internal states of the first agent and that despite its going through a second agent's intentional action is nonetheless an intentional action of the first agent. I take it that strong intuitions play against accepting this alternative.

causing such behaviour also rationalize it.[5] A crucial aspect of this way of understanding agency is that it sees the relevant agential systems as dynamic centres of behavioural control extending their agential reach through functional mechanisms.

Consequently, Bishop's general take on causal deviance is that what distinguishes deviant from non-deviant causal chains depends on the sort of behavioural control that agents normally exercise and that is lost whenever deviance occurs. This distinction is present in deviant cases involving prosthetic agents:

> Sometimes the second agent's involvement in the causal chain *preempts or blocks the agent's exercise of direct control* over his or her bodily movements. Then the second agent is no mere cog in the mechanisms that realize the first agent's direct control. Rather, the second agent is part of a system that provides the first agent with, at best, only *indirect* control over the movements of his or her own body. [1989: 159]

Thus, in order to distinguish deviant from non-deviant cases it is necessary to identify the specific features of the second agent's intervention capable of undermining the first agent's control over her bodily movements. Bishop identifies these disruptive features by focusing on the "sustained" nature of most intentional actions. [6]

An intentional action is sustained when it involves a process whereby its agent regulates its execution by monitoring whether the ensuing behaviour has been completed or whether it has satisfied the intended goal. This process necessarily involves feedback information arising from the behaviour under execution. Bishop proposes that under normal conditions the information is fed to the brain of the subject who triggers the chain of events leading to the behaviour. However, under abnormal conditions the information may never reach the subject's brain. Bishop contends that this last thing is exactly what happens in deviant cases when a second agent disrupts the control of the first agent over her bodily movements, for in such cases:

> There is a servo-system functioning to match the agent's intention all right; but given its detailed architecture, it can hardly count as realizing *the agent's* controlled regulation of his or her bodily movements since the *feedback information* about orientation and muscular states does not get carried back to the agent's central processing system. [1989: 170]

To illustrate Bishop's proposal let us try it on Peacocke's neurophysiologist case. If indeed the neurophysiologist is capable of reading the behavioural intentions from a patient's brain in order to produce behaviour that exactly matches their content, then the feedback information involved in the execution of this behaviour will miss the patient's brain. That is, if anyone is getting this feedback information this person would be the neurophysiologist. This results in the patient's lack of behavioural control revealing its deviant nature. In contrast, when the assistant simply holds the wire nothing prevents the feedback information from reaching the patient's brain allowing her to control her bodily movement. This explains why the assistant's case is not deviant.

---

[5] Other versions inspired on similar systemic approaches to action and agency include: [Enç and Adams: 1992; Jeannerod: 1997; Clark: 1997; Juarrero: 1999; and Enç: 2003].

[6] The inclusion of such sustained character in the production of an intentional action goes back at least to Irving Thalberg [1984] who in turn was inspired by a critic of CTA, namely, Harry Frankfurt [1978]. The extensive literature on guiding intentions directly captures some features of such sustained aspect of many intentional actions. See, for example, [Searle: 1983] and [Mele: 1992].

Bishop then proposes the inclusion of feedback as a necessary condition to identify an event as a basic intentional action. In fact, he believes that the addition of this condition completes the set of necessary and sufficient conditions to identify a basic intentional action. Hence, according to Bishop:

> *M* performs the basic intentional action of *a*-ing if and only if,
> (1) *M* has a basic intention to do *a*; and,
> (2) *M*'s having this basic intention causes *M* to produce behaviour, *b*, which instantiates the types of state or event intrinsic to the action of *a*-ing; where the causal mechanism from *M*'s basic intention to *b* satisfies the sensitivity condition; and if this causal mechanism involves feedback, then the feedback signal is routed back to *M*'s central mental processes if to anyone's. [1989: 172]

Thus, besides sensitivity which Bishop requires to preclude deviance in typical causal chains, his proposal adds the feedback condition to deal with the peculiar deviance that emerges in cases involving prosthetic agents. This move involves the recognition of a feedback signal as a constitutive element of most intentional actions. That is, the relevant causal chain of events producing an intentional action involves in most cases a two-way causal sequence of events: One that goes from the agent's mental processes to the bodily movement and another one that sends back the information to the agent's mental processes about the bodily movement.

Problems with Bishop's Move

Unfortunately, Bishop's reply is insufficient to deal with deviant cases involving prosthetic agents like Peacocke's neurophysiologist. Even the addition of the feedback condition to the sensitivity condition does not eliminate the possibility of having cases where these conditions are met and yet deviance occurs.

Let us remember that in non-deviant prosthetic cases Bishop is willing to grant that a second agent's intentional action can be placed between the intentional states of the first agent and her bodily movements without this undermining her agential control. That is, Bishop accepts that cases involving prosthetic agents by themselves are not the source of deviance. Rather, he proposes that deviance occurs only when the second agent blocks the control of the first agent by stopping the feedback information from reaching the first agent. However, let us note that in order for this proposal to even start making sense the causal chain going from the first agent to the movement of her body needs to go through the second agent's intentional action. This in itself does not yet distinguish this causal chain from a non-deviant one, for it is only in the process of receiving information from the bodily movement that the wanted distinction supposedly arises. Bishop then asks whose brain is getting the feedback information: If it is the second agent, then we are dealing with a deviant case. But, if it is the first agent who is getting the information, then this is not a deviant case.

The problem with this suggestion is that from a systemic approach to CTA nothing prevents enriching the neurophysiologist scenario with the possibility of a further link now going from the neurophysiologist to the patient's brain. This link will pass the information received from the bodily movement to the patient's brain in a similar way in which the neurophysiologist got the information from the patient's intentional states. If causal transitivity was sufficient to go in one direction, causal transitivity should be sufficient to permit the flow of information to go in the other direction. If this occurs, then both brains receive feedback information and hence this sole feature cannot be what distinguishes cases that are deviant from cases that are not deviant.

Bishop can answer this objection by rejecting the capacity of a second agent to send back the information to the first agent. But, this answer is clearly at odds with his acceptance of non-deviant cases like the assistant's. If we can conceive that the assistant is capable of bridging the patient's brain with the patient's behaviour, then we can conceive that the assistant is capable of bridging the patient's behaviour with the patient's brain. Analogously, if we can conceive that the neurophysiologist is capable of bridging the patient's brain with the patient's behaviour, then we can conceive that the neurophysiologist is capable of bridging the patient's behaviour with the patient's brain. Why should the causal role of these intermediate agents be different when the information goes in different directions? Alternatively, Bishop can insist that his feedback condition explicitly says that if this causal mechanism involves feedback, then the feedback signal is routed back to the first agent's central mental processes if to anyone's. But, this is precisely what is at stake in this criticism.

In fact, this alternative possibility where the neurophysiologist acts as a functional bridge of feedback information is completely consistent with a systemic approach to agency. For what emerges in this enriched scenario is a larger agential system that makes use of two agents. And, once we see the two agents as subsystems of a larger agential system, Bishop's proposal runs into the mentioned problem. Moreover, by expanding the agential system to include alternative functional bridges of information another difficulty for Bishop's proposal becomes apparent.

Presumably, the neurophysiologist scenario does not involve the possibility of obtaining feedback information through some external way, say, by the patient observing whether the neurophysiologist is satisfying her bodily intentions. Nevertheless, a systemic account of agency like Bishop's is fully compatible with the possibility of external feedback informing the execution of an intentional action. However, once the possibility of external feedback is introduced, the way in which Bishop proposes to deal with deviance runs again into difficulties. Now it is possible to imagine a scenario where the external feedback information is obtained through a deployment of mirrors and fake images.[7] Such arrangement would offer visual information that the agent can use to try to control some of her bodily movements. So, if the agent ever succeeds in moving her body in the way in which she intends to move it—and nothing seems to preclude this possibility—a counterexample to Bishop's conditions is produced. That is, the agent has blindly produced a bodily movement that only accidentally matches her bodily intention despite the fact that: (1) it is her intention that causes the bodily movement; (2) the causal mechanism involved in this movement satisfies the sensitivity condition; and (3) the relevant feedback information is going back to her "central mental processes if to anyone's."

The only way in which Bishop can apparently avoid this type of deviant scenario seems to be by strengthening the feedback condition to require that the source of information must originate in the relevant bodily movement. Not only is such strengthened condition not present in Bishop's proposal but its articulation involves facing well known problems with epistemic deviance and semantic information. [8]

Finally, there is a difficulty with Bishop's proposal concerning not so much the overall architecture involved in systemic agency but the specific nature of the relevant internal states

[7] Similar scenarios are often employed in empirical research concerning agency. Perhaps the most famous example is the one carried out by Nielsen [1963] 'where subjects where unknowingly shown an alien hand in exact concordance with their own hand' producing the strong belief that it was their own hand that they were moving and guiding. [Jeannerod 1997: 186].

[8] For this problem to arise we do not need to bring in the second agent's contribution. Moreover, since feedback is essentially a cognitive feature of the agential system its presence opens the door to relevant considerations with respect to deviance coming from discussions in epistemology.

guiding the execution of bona fide intentional actions. A main virtue of approaching agency from a systemic perspective is that it nicely incorporates the dynamic nature of agency. Hence, a critical component of this effort involves the introduction of the relevant internal states that can play the required role to account for such dynamic agency. In the case of Bishop basic intentions are supposed to play this critical role. However, as soon as we consider the nature of such basic intentions a rather pressing question immediately arises, namely, are the same basic intentions that start the causing of bodily movements the ones that are involved in their execution? If they are the same basic intentions, then Bishop needs to defend the idea that such intentions are continuous. If they are not the same basic intentions, then he needs to defend the idea that there is some way in which different basic intentions are able to work in tandem and respond to feedback information. Although it seems that the most plausible proposal is the first one, as it stands Bishop's position is silent concerning this issue.[9]

In conclusion, a systemic approach to agency remains one of the most attractive features of contemporary versions of CTA. Nevertheless, such approach also continues to face a very serious challenge coming from prosthetic deviance. If this is correct, a "Final Breakthrough" concerning the necessary and sufficient conditions for intentional action remains a tantalizing but as yet unrealised theoretical possibility.

References

Bishop, John 1989. *Natural Agency: An Essay on the Causal Theory of Action*, Cambridge: Cambridge University Press.

Bishop, John 1997. Naturalising Mental Action, *Contemporary Action Theory. Vol. 1* (Eds. G. Holmstrom-Hintikka and R. Toumela), Dordrecht: Kluwer Academic Press.

Clark, Andy 1997. *Being There: Putting Brain, Body, and World Together Again*, Cambridge, Mass.: MIT Press.

[9] The key question here is how an internal state is capable to be of the required sustaining nature to count as "the same" intention and how this intention is supposed to monitor the ensuing behaviour. One recent discussion of the type of mental states involved in such sustaining and monitoring task has been offered by François Schroeter [2004]. Bishop partially addressed this difficulty in [1997], but as he has recognized in personal communications this is still a pending work in his proposal.

Enç, Berent and Fred Adams 1992. Functions and Goal-Directedness, *Philosophy of Science* 59: 635-654.

Enç, Berent 2003. *How We Act: Causes, Reasons, and Intentions*, Oxford: Clarendon Press.

Frankfurt, Harry 1978. The Problem of Action, *American Philosophical Quarterly* 15: 157-62.

Hart, H. L. A. and A. Honoré 1959. *Causation and the Law*, Oxford: Oxford University Press.

Jeannerod, Marc 1997. *The Cognitive Neuroscience of Action*, Cambridge: Blackwell.

Juarrero, Alicia 1999. *Dynamics in Action: Intentional Behavior as a Complex System*, Cambridge, Mass.: MIT Press.

Mele, Alfred 1992. *Springs of Action*, New York: Oxford University Press.

Nielsen, T.I. 1963. Volition: A new experimental approach, *Scandinavian Journal of Psychology* 4: 225-30.

Peacocke, Christopher 1979. *Holistic Explanation: Action, Space, Interpretation*, New York: Oxford University Press.

Schroeter, François 2004. Endorsement and Autonomous Agency, *Philosophy and Phenomenological Research* 69: 633-59.

Searle, John 1983. *Intentionality*, Cambridge: Cambridge University Press.

Thalberg, Irving 1984. Do Our Intentions Cause Our Intentional Actions? *American Philosophical Quarterly* 21: 249-60.