

PAUL NOORDHOF

ENVIRONMENT-DEPENDENT CONTENT AND THE VIRTUES OF CAUSAL EXPLANATION

One line of objection to Externalism or Anti-Individualism is that, if it were true, causal explanations citing the content of mental states or events would be redundant. Another is that individuation by intentional properties would fail to be individuation by causal powers. My aim in the present paper is to answer these charges. But first a few terminological matters to get the charges into focus.

We need a formulation of Externalism or Anti-Individualism. Let us say that:

A token mental state (or event) m 's possession of the content that p by a subject S is environment-dependent iff it is not metaphysically possible for m to possess the content that p given no change in S 's remaining total mental state (i.e. the subject's total mental state excluding m) unless the socio-linguistic and/or natural environment is or was or will be a certain way.

Obviously the characterisation of environment can be broadened to include, for instance, the home of numbers and universals. This would be necessary in order to capture Colin McGinn's doctrine of Weak Externalism (McGinn 1989, 7). However, nothing in the present discussion hangs upon it. The formulation in terms of token mental states and its relativisation to S 's remaining total mental state is to deal with the following case. Suppose that my thought that water is wet is environment-dependent. Yet if I knew more about chemistry, and were more inventive, I could define a substance uninstantiated in my world as having the molecular structure H_2O . In such a changed total mental state, the same thought would not be environment-dependent. But this does nothing to undermine the environment-dependence of my thought about water given how I actually am. With this understanding of when a token mental state or event is environment-dependent, we can define Externalism or Anti-Individualism as follows: some mental states or events possess contents environment-dependently (for defence of such a claim, see Burge 1979, 1986; Evans 1981b, Ch. 9).

Contents are truth conditions, for instance, objects and properties or states of affairs or sets of possible worlds. I shall not seek to come to a definite view on this. Causal explanations of behaviour citing content-bearing mental states and events (such as beliefs, desires, and so on) do not primarily (if at all) record the efficacy of truth conditions. The threat to Externalism does not derive from concern over whether sets of possible worlds, for instance, are efficacious. Instead, citing the content of a mental state or event is a way of picking out a property which does seem to have a causal explanatory impact upon behaviour, namely the intentional properties of the mental states and events. Intentional properties are those properties of mental events or states *in virtue of which* they have contents. Why did Jo keep his hand away from the goose? Because he believed *that the goose might bite him*. The intentional property of Jo's state of belief explains why it is appropriate to attribute to him the content that the goose might bite him and this intentional property seems to have an impact upon his behaviour. Again, nothing hangs upon the particular way in which we carve up the problem but the distinction just made needs to be recognised somehow or another. If Externalism is true, then some intentional properties involve relations to the environment. It is from this fact that problems are thought to arise.

Sometimes the challenge is put in terms of the concern that the intentional properties of mental states and events will fail to be efficacious or causally relevant if Externalism is true. The worry is that the intentional property in virtue of which Jo is ascribed the belief that the goose will bite him has no impact on his behaviour (contrary to appearances). However, this does not seem the best way to put the issue. I shall assume, and I have argued elsewhere, that intentional properties are efficacious because they are partly constituted from internal (i.e. non-environment-dependent) properties of subjects (Noordhof 1999b; this paper partly develops a claim in Segal and Sober 1991). A simple analogy conveys the thought. If I take a hammer to a statue and smash the statue, we rightly say that the hammer blow was a cause of the smashing of the statue. Moreover, we take ourselves to have picked out the event – the hammer blow – by a causally relevant description. Although less idiomatic, it is perfectly correct to say that it was the object's property of being a hammer which was a cause of the smashing of the statue. It was the instancing of hammerness travelling in a certain fashion which constituted the event which caused the

smashing. Nevertheless, the hammer has lots of relational properties, which must hold if the hammer is to be what it is: a *tool* and not just an object of a certain mass and resistance. We don't take the partly relational character of the property of being a hammer to undermine its causal status even though the smashing is a result of the mass and resistance of the hammer. I urge that the situation is the same with regard to intentional properties in the case of Externalism being true.

Some argue that Externalism is committed to the claim that intentional properties are wholly relational (Dretske 1996). But that is not right. Three plausible considerations indicate that intentional properties are partly constituted by internal properties. First, the relational properties constitutive of intentional properties could not be instantiated unless there were instances of internal properties to stand in the relevant relations. For instance, if you suppose that intentional properties are a matter of causal covariance then internal properties must covary with properties in the environment. Second, the intentional properties of propositional attitudes – unlike intentional properties of sub-personal states perhaps – ensure that the contents of the relevant mental states and events are *for* the subjects in these states or undergoing these events. At the minimum, this means that the contents of mental states and events are well integrated into a subject's mental life. The required integration implies that subjects of these states (or undergoing these events) have a certain internal structure which, hence, partly constitute the intentional properties of the states or events. Third, and finally, different intentional properties correspond to different modes of presentation of the objects and properties thought about. These differences in modes of presentation will once more be partly an internal matter, for instance, captured in terms of the distinctive functional roles of the mental states and events with these intentional properties.

Accepting that internal properties partly constitute intentional properties (whether or not Externalism is true) does not enable the Externalist to avoid the challenge to the causal explanatory role of intentional properties. It just means that the challenge should not be formulated in terms of a lack of efficacy. Instead, as I indicated at the outset, it should be formulated in terms of concern over causal explanatory redundancy and failure to correspond to differences in causal powers. These are the focus of my discussion. It proceeds as follows. In the first section, I will formulate what I call the *Argument from Precision* which explains how intentional properties may

appear to be causally explanatorily redundant on the Externalist picture. I will then turn to discuss the *Argument from Causal Powers*. This is to the effect that intentional properties should not figure in scientifically advanced causal explanations if Externalism were true because individuation by intentional properties is not individuation by causal powers. I will explain why this argument is defective and argue that, once we recognise that this is so, we will be able to resist the *Argument from Precision*. The paper closes with a discussion of two virtues of causal explanation. For those who do not like any suspense, let me just say that my line will be that it is because individuation by intentional properties is individuation by psychologically important causal powers that reference to content-bearing mental states and events fails to be redundant in causal explanations of behaviour.

1. THE ARGUMENT FROM PRECISION

The Argument from Precision is meant to establish the following conclusion:

If a property F is a cause of G in circumstances C just because it is partly constituted from something which is a cause of G in C, then reference to e_1 having F in a causal explanation of e_2 having G is redundant.

The obvious implication for mental states or events with environment-dependent content is that reference to the intentional properties in virtue of which they have such contents will be causally explanatorily redundant even if these properties are efficacious. Concern over when it is appropriate to say that some properties constitute other properties and what it means to say that properties are efficacious requires more precise formulation of these issues.

The first point to make is that my talk of one property causing another up until now has just meant that an instance of the first property causes an instance of the second. Although I think it is possible to specify a notion of property causation, and not just property instance causation, this will not be my concern in what follows (see Noordhof 1999b). Similarly, my talk of properties being partly constituted from other properties is shorthand for the idea that instances of properties may be constituted partly or wholly from instances of other properties. I am not claiming that

properties constitute other properties if this is taken to be different from constitution of instances. There are well known difficulties with the idea of properties constituting other properties if the focus is not on instances of properties but the properties themselves (as it were). For example, the property of being H_2O would presumably have to be constituted from the properties of hydrogen and oxygen. But are there enough properties of hydrogen? Intuitively there is one property of hydrogen instantiated by many atoms. However, the property of being H_2O would seem to require two. If two or more properties of hydrogen are allowed to exist, what do they have in common? The answer seems embarrassing (Lewis 1986, 31–46).

These qualifications made, my proposal is this:

An instance of an A-property, P, causes an instance of an A-property, Q, if and only if

- (i) An instance of a B-property, K, is part of the minimal supervenience-base of the instantiation of P and an instance of a B-property, J, is part of the minimal supervenience-base of the instantiation of Q.
- (ii) The instantiation of K causes the instantiation of J (cf. Kim 1984; Segal and Sober 1991; Noordhof 1999b).

Supervenience comes in various forms. My proposal does not hold good for all of them. I think it does hold good for the following kind (a kind of strong supervenience in Kim's sense).

A *supervenies* on B just in case, necessarily_n, for each x and each property F in A, if x has F, then there is a property G in B such that x has G, and *necessarily*_m if any y has G, it has F. i.e. $\Box_n (x)(F)(Fx \ \& \ F \ e \ A \ \rightarrow \ (\exists G) (G \ e \ B \ \& \ Gx \ \& \ \Box_m (y)(Gy \ \rightarrow \ Fy))$ (cf. Kim 1984, p. 65).

(where A and B are families of properties, the supervening and supervenience-base (or subvenient) properties respectively, “e” is “is a member of”, and ‘ \Box ’ is the necessity operator with ‘ \Box_n ’ meaning nomological and ‘ \Box_m ’ metaphysical necessity respectively).

The first modal operator is taken to be nomological necessity to leave it open that some supervening properties may contingently supervene on their supervenience-base. In very different worlds with very different laws, there may be entirely different supervenience-bases to those of type B. For instance, the property of being a hammer may supervene on objects without mass in some weird and wonderful world with very different laws. Taking the first operator to be nomological necessity captures the fact that the instantiation of the property of being a hammer requires the instantiation of the property of mass in worlds like ours while leaving open the more exotic possibilities mentioned. The interpretation of the second

modal operator as metaphysical necessity seems appropriate because I am trying to capture the idea that instances of the supervening properties are in some way constituted from instances of their supervenience-base properties – as suggested by the examples I have discussed. The notion of constitution does not involve something as weak as mere nomological correlation between F and G, metaphysical necessity seems to fit the bill.

Although, supervenience so understood goes some way to capture the idea of constitution, it is not sufficient by itself. Suppose that the property of being a charged battery supervenes in the required sense upon certain electro-chemical properties. Then it also supervenes on certain electro-chemical properties and the property of being a raised flag. Yet the battery is not partly constituted from being a raised flag. This would not matter much if it had no consequences for what we might be inclined to say about the efficacy of batteries. But if we are seeking to explain how the efficacy of a property is affected by the efficacy of the properties which constitute its instantiation, then we cannot afford this result. The property of being a raised flag caused the military man to salute. The presence of the battery did not. The causal contribution of the property of being a raised flag has nothing to do with the causal contribution of the battery. So we had better not adopt an understanding of constitution that suggests otherwise.

We need the idea of a minimal supervenience-base of a property understood as follows.

G is part of a minimal supervenience-base B of F if and only if

- (a) G is a member of B.
- (b) B is a set of properties {G, H, I, ...} such that metaphysically necessarily, if all the members of B are coinstantiated, then F is instantiated.
- (c) It is not the case that metaphysically necessarily if all the members of B except G are instantiated, then F is instantiated.

The property of being a raised flag would not count as part of the minimal supervenience-base of the property of being a charged battery since, if the property of being a raised flag were deleted from the set of properties which metaphysically necessitate the instantiation of the property of being a charged battery, the remaining set would still metaphysically necessitate the instantiation. With this in place, I think we have some idea of how an instance of the property of being a hammer, to return to our initial example,

can be efficacious because it is partly constituted from an instance of the property of being an object with a certain mass and resistance.

We can now formulate the conclusion of the Argument from Precision as follows.

If a property F is a cause of G in circumstances C just because a component of its minimal supervenience-base, is a cause of G in C , then reference to e_1 having F in a causal explanation of e_2 having G is redundant.

The argument for the conclusion runs as follows.

- (1) If a property F is a cause of G in the circumstances C just because a component of its minimal supervenience-base, is a cause of G in C , then we could explain e_2 having G in circumstances C just by mentioning the component of the minimal supervenience-base of F .
- (2) If we could explain e_2 having G in circumstances C just by mentioning the component of the minimal supervenience-base of F , then reference to e_1 having F in a causal explanation of e_2 having G is redundant.

Therefore,

- (3) Reference to e_1 having F in a causal explanation of e_2 having G is redundant.

Applied to my running example, the property of being a hammer can be cited to explain the smashing of a nut (with the hammer). However, the explanation is redundant. Talking about an object with such and such a resistance and mass would be better. Applied to a more contentious case of central interest for us, an environment-dependent intentional property – that water_{H₂O} is wet – can be cited to explain why a subject took an umbrella when it was raining. However, the explanation is causally explanatorily redundant because what is common to this belief and belief that water_{XYZ} is wet will explain why a subject took an umbrella.

Note I am not making the claim that such an explanation would be redundant tout court (if the argument were sound). It might capture a rationalisation of a subject's behaviour that would not be available by focussing on what is common to these two beliefs because there is no relevant common content. I don't want to rule this out. I'm concerned with the claim of explanatory redundancy solely as a claim about the causal worth of the explanation.

The argument brings out the fact that one virtue of causal explanation is *precision*. We can refine the causal explanations we offer by providing more precise characterisations of what is efficacious. If there is a component of the minimal supervenience base of F which is a cause of G, and no other component is, then it would be better merely to cite that. However, premise (2) of the argument is true only so long as precision is the sole relevant virtue of causal explanation. If there are other relevant virtues then the availability of another explanation citing part of the minimal supervenience-base of environment-dependent intentional properties does not imply redundancy. It is this that I will seek to establish by turning to the *Argument from Causal Powers*.

2. THE ARGUMENT FROM CAUSAL POWERS

At first sight, things don't look promising. According to the Argument from Causal Powers, it seems that explanations involving intentional properties lack another virtue of causal explanations. The argument runs as follows:

- (1) Two different intentional properties may be attributed to internally identical subjects.
- (2) If two different intentional properties may be attributed to internally identical subjects, then these intentional properties have precisely the same causal powers.
- (3) Intentional properties are not discriminable by their causal powers (from (1) and (2)).
- (4) Scientifically advanced causal explanations should only involve properties discriminable by their causal powers.

Therefore,

- (5) Intentional properties should not figure in scientifically advanced causal explanations.

Jerry Fodor is responsible for the formulation of this argument. His actual formulation makes the argument explicitly an argument for Internalism (the denial of Externalism) (Fodor 1991, 5, 25). I have just drawn out that bit of it which presents the challenge to Externalism.

Key to a proper assessment of the argument is the idea of two properties being discriminable by their causal powers or, as Fodor sometimes says, individuated in terms of their causal powers. This is not a commitment to a causal theory of properties. It is not being claimed that properties have their causal powers essentially and that, necessarily, a property would have the same causal powers in every possible world (see e.g. Shoemaker 1980). I don't mean to rule this out. Just to indicate that such a claim is not in play. In which case, talk of individuation is not to be understood in terms of specifying the essential nature of a property. Nor is it quite to be understood in terms of individuation of objects in terms of their spatiotemporal location. Here contingent features, those relating to objects' precise spatiotemporal location, are plausibly thought to be needed to discriminate between objects. Instead, talk of individuation just seems to refer to an important way in which two properties can be distinguished from each other even if they might also be discriminable in other ways.

The characterisation that proponents of the argument seem to have in mind is as follows.

(CP) P_1 and P_2 have distinct causal powers iff either (i) for some E , there are some nomologically possible circumstances C , in which P_1 causes E and P_2 does not or (ii) one of P_1 and P_2 is not instantiable in any nomologically possible world and the other has causal powers in the actual world.¹

A nomologically possible world is a world with the same laws as those of the actual world. Properties instantiated in the actual world will have different causal powers to those which cannot be instantiated in any nomologically possible world because, assuming that properties instantiated in the actual world have causal powers at all, they will have causal powers which the other properties lack. However, this is not the case which proponents of the argument from causal powers have in mind. It is just mentioned to complete the analysis of distinctness of causal powers. Their focus is on the first possibility. If two properties have distinct causal powers, then there should be some circumstances in which one has an effect that the other would not have: the 'distinctive causal contribution'. I note, because this will be relevant shortly, that this is quite compatible with them having the same causal powers in other circumstances.

Applying the argument to a particular case, the familiar thought is that two internally identical subjects, Fred and twin Fred in different environments, one containing water_{H₂O} and one containing

water_{XYZ}, will have different beliefs and desires in virtue of having different contents concerning water_{H₂O} and water_{XYZ}, respectively, and yet their behaviour will be precisely the same. Hence the difference between water_{H₂O} and water_{XYZ} intentional properties is not a difference of causal powers.

There have been a number of responses to the argument. Some have argued that sciences don't necessarily focus on properties discriminable by their causal powers and hence reject premise (4) of the argument. For instance, they note that biological species are classified by causal historical means. At the extreme, two animals may belong to different species if they are reproductively isolated and phylogenetically independent even if they are biochemically identical (Wilson 1992, 1995, 44). In reply, sceptics might agree that there are lots of causal explanations involving properties which are not discriminable by their causal powers. However, they will claim, this doesn't answer the suspicion that these causal explanations lack virtue. If we focussed on causal explanatory virtues alone, they would be replaced with other explanations. At best they would have a derivative virtue. It would be a merit in explanations citing scientifically important properties which are not discriminable by their causal powers, that the causal explanations mention causal relations in which they stand. But the merit would derive from the fact that there is independent reason to want to classify things in this other way. It is not how one would choose to classify things if one were interested in the character of causal explanations alone.

Another reply is that it is distinctive of the special sciences that the properties to which they refer may be characterised by the causal powers they would have if instantiated *in their normal environment*. For example, the property of being a human heart is characterised by its causal powers in the human body. Being merely a certain kind of pumping mechanism does not have the human body as its normal environment (Burge 1989, 309–311). So even if a pumping mechanism of a certain kind would not necessarily be a heart and yet have the causal powers of a heart, it would not follow that we should refer to pumping mechanisms of appropriate kinds rather than hearts in our special science explanations.

To this suggestion, I have two concerns. The first is that it is not clear that appealing to this weaker notion makes things better. The normal environment in the case of intentional properties is minds finding out information and reasoning about the world. It is precisely in this environment that the charge is made that,

if Externalism is true, differences in intentional properties are not differences in causal powers. There are no grounds for adopting a more restricted normal environment in seeking to characterise the causal powers of intentional properties. It is no part of the claims concerning Fred and his twin that they are unable travel to each other's environment and interact with it. Nor, as I indicated at the outset, is it ruled out that an inventive and knowledgeable subject may think about XYZ without being in an XYZ environment. The second concern is that it would not be unreasonable for special scientists – once the issue about causal powers is on the table – to provide explanations which identify entities which differ by their causal powers more substantially understood. The blood is pumped round the body in such and such a way because the body contains a pump mechanism of such and such a kind (which we call a heart).

Both of the two responses sketched above, in effect, reject premise (4) of Fodor's argument. By contrast, I will argue that, even if Externalism is true, individuation by intentional properties is individuation by causal powers. The problem is that the argument from (1) and (2) to (3) is unsound. This is disguised by the ambiguity in premise (2).

(2) If two different intentional properties may be attributed to internally identical subjects, then these intentional properties have precisely the same causal powers.

If (2) means

(2A) If two different intentional properties may be attributed to internally identical subjects, then these intentional properties have precisely the same causal powers *in these subjects at the present time*.

then (2) is true. But (3) doesn't follow because (2A) does not assert that the intentional properties' causal powers will be the same in all nomologically possible situations. If (2) means

(2B) If two different intentional properties may be attributed to internally identical subjects, then these intentional properties have precisely the same causal powers in all nomologically possible situations.

then (2B) is false, or so I seek to establish.

In order successfully to carry out this strategy, certain tempting lines of response need to be avoided. It is clearly pretty easy to establish that environment-dependent intentional properties have some kind of distinct causal powers. For instance, suppose that I know that Fred has come from Earth and twin Fred from Twin

Earth. Then Fred's desire that he have a glass of water expressed by 'Please give me a glass of water' and twin Fred's desire expressed by 'Please give me a glass of water' will have different causal consequences on me. I will bring either H_2O or XYZ (Burge 1989, 311–313). However, we are not looking for any old distinct causal contribution but distinct causal contributions regarding the types of entity which are the typical explananda for distinctions between content-bearing states viz. the behaviour of the subjects to whom intentional properties are ascribed.

Another tempting line of response which should be resisted is what I shall call the Broad Behaviour Strategy. The Strategy insists that differences in the causal powers of intentional properties show up once we recognise that the behaviour of the subjects on earth and twin earth is, in fact, different. There are various versions of the strategy depending upon the kind of individuation of behaviour adopted but two should serve to indicate the problems with the strategy. According to one version:

(B1) Behaviour of kind B could only occur if the objects and properties to which reference is made in the standard description of a piece of behaviour of this kind exist at the time of the behaviour (Peacocke 1981, 1993, 210; Hornsby 1986, 123–126).

A standard description of a piece of behaviour is Fred *reached out for a drink of water* H_2O . According to (B1), Fred could not reach for a drink of water H_2O unless there were H_2O in his environment. Similarly Fred's twin could not reach for a drink of water XYZ unless there were XYZ in his environment. Fred's intention to reach for water H_2O would, thus, give rise to water H_2O -reaching-behaviour whereas Fred's twin's intention to reach for water XYZ would give rise to water XYZ -reaching-behaviour.

The problem with this suggestion is that it is implausible that Fred's intention to reach for water H_2O is causally responsible for the distinctive feature of his water H_2O -directed-behaviour, namely that it is behaviour directed towards water H_2O actually present in Fred's environment. Rather it is the presence of water H_2O in the environment which ensures that his behaviour is water H_2O -directed as opposed to water XYZ -directed. For instance, Fred's intention to reach for water H_2O does not raise the chance of water H_2O existing in his immediate environment (given that the worldly conditions ensure that it is there).

The alternative version of the Broad Behaviour Strategy individuates behaviour as follows.

(B2) Behaviour of kind B could only occur if the subject were to have an intention whose content matches the standard description of the behaviour (Fodor 1991, 8–9, 21, for discussion but not endorsement).

According to (B2), Fred could not reach for a drink of water_{H₂O} unless he had the intention to reach for water_{H₂O}. Similarly Fred's twin could not reach for a drink of water_{XYZ} unless he had the intention to reach for water_{XYZ}. Now it is clear that the respective intentions do have a distinctive contribution to make regarding which behaviour occurs. However, there are two hypotheses about the basis for this distinctive contribution.

(H1) The way in which the behaviour is individuated is solely responsible for the difference in behaviour, the causal contribution of the two different intentions is the same.

(H2) Although the way in which the behaviour is individuated implies that a certain kind of intention is logically necessary for the behaviour, each intention also makes a distinctive causal contribution to the behaviour.

Those who question whether environment-dependent intentional properties make a distinctive causal contribution are not going to be satisfied with (H2) when (H1) explains the difference in contribution just as well. We need to have some reason to suppose that (H2) is correct. That has not been provided by the Broad Behaviour Strategy.

Instead, my proposal is that the distinctive causal contribution of states with intentional properties partly specified in terms of natural kind K (say) is revealed in the minds of *experts* about K. Experts regarding a particular natural kind K are those who can distinguish K from all others, even those which are apparently similar. They need not be experts concerning every kind because they may be unable to distinguish any other kind from every other one. They could just be experts about water or gold (say). It is within the minds of experts about a natural kind K (hereafter 'K-experts') that a K-intentional property (i.e. one in virtue of which contents of the form *that K is ...* or *that ... is K* are ascribed) will make its distinctive appropriate causal contribution, standardly, by giving rise to bodily movements that distinguish K from other kinds. For instance, XYZ-experts will not go swimming in the Thames if they want to go swimming in XYZ. The bodily movement in question is not to be individuated environment-dependently. The point is rather that,

if the expert is on a diving board by the Thames, the bodily movement (e.g. diving) will fail to occur if they are an XYZ-expert and wish to go swimming in XYZ whereas it would occur if they were an H₂O expert and had the corresponding wish to go swimming in H₂O. Given that K-experts are nomically possible, (2B) is false. XYZ and H₂O intentional properties have different causal powers because they have a distinctive causal contribution in some nomically possible circumstances. Hence the argument in favour of (3) is unsound.

The mistake has been to suppose that the distinctive causal contribution should be revealed in twin subjects. Since psychology is in the business of trying to explain the behaviour of expert and nonexpert subjects, it is no surprise that it should adopt a means of individuating psychological properties which need only display distinct causal contributions in the expert. A consequence of my response is that the intentional properties have causal powers which outstrip those of a subject's internal properties. However, this should also come as no surprise. The supervenience-bases of intentional properties are not just internal properties.² This may be so even in the minds of the expert. Burge has suggested that experts could not have the appropriate water_{H₂O}-content unless embedded in the right language community (Burge 1979, 84–85). Indeed, his thought seems to be that somebody is a water_{H₂O} expert (rather than an ignoramus about water_{XYZ}) only if his or her speech community possess a term for water_{H₂O}. In which case, it will be an environment-dependent matter that a particular internal property which displays the appropriate distinctive causal contribution for water_{H₂O} is part of the supervenience-base for the intentional property relating to water_{H₂O}. Only if the relevant experts are not deferential, and so not seeking to have their concept of water correspond to some linguistically available concept of water or some kind present in their environment, does it seem that their possession of water-contents is environment-independent.

One objection to the proposal is that, while I might have explained how environment-dependent properties have a distinctive causal explanatory role in the minds of the experts, I have failed to justify their attribution to explain the behaviour of the inexpert. There are three points to make in reply. First, all I have tried to establish at the moment is that individuation by intentional properties is individuation by causal powers concerning the typical explananda of attributions of intentional properties (against Fodor's

argument). Second, the explanatory worth of the attributions of environment-dependent intentional properties to nonexpert subjects relies on the explanatory worth of explanations citing properties individuated in terms of causal powers. I will say more about this in the next section but for now it is just worth noting that the background to the objection seems to be an assumption that precision is the only virtue of causal explanation. Third, and finally, the inexpert may become expert. While the attribution of a certain intentional property to an inexpert subject may not have a distinctive causal impact now, as a subject's expertise grows, he or she will be more sensitive to the fulfilment conditions of the intentional property. So the attribution has a potential diachronic explanatory worth which justifies the present attribution.

Another objection to my proposal is that even in the minds of the expert, the distinct causal contribution can be explained in terms of environment-independent properties e.g. neural properties of the brain or environment-independent intentional properties (those in virtue of which the fabled narrow content is attributed). I don't dispute that this is so but it doesn't impugn the point I've made regarding the distinctive causal powers of intentional properties. The environment-independent properties will be part of the supervenience-base of intentional properties and what efficacy the former properties have will be inherited by the latter (for further discussion, see Noordhof 1999b). Intentional properties are realised in a multi-fold of different ways and it is this multiplicity that provides them with causal powers not shared by any of their realisers.

Perhaps the most pressing objection to my proposal concerns whether experts of the sort I have identified are possible. The worry can be expressed in different ways. It might be argued that, if somebody is an expert, then there is a metaphysically necessary relationship between certain external differences in the environment and certain internal differences in the expert. But no such relationship could exist. It is always metaphysically possible for the appropriate internal properties to be instantiated without the items in the environment or vice versa.

This way of putting the concern appears misguided. The objector is mistaken in assuming that I must postulate a metaphysically necessary relationship between external items and internal properties of the agent. Rather the commitment is appropriately described as follows:

Metaphysically necessarily, if *S* is a *K*-expert, then, in normal circumstances, if *K* is present in the environment and *S* seeks to establish whether there is *K* in the environment, then *S* will believe that there is *K* and will respond differently in virtue of having this belief than any other belief, should this be appropriate.

This is quite compatible with the connection between external items and internal properties of the expert breaking down in abnormal circumstances, or even in normal circumstances if the expert ceases to be an expert. Indeed, the formulation just given may be too strong. Even experts can make mistakes in normal circumstances just so long as they don't do it enough to impugn their expertise. Differences between intentional properties that inexpert twins cannot distinguish would still have a distinctive causal contribution to make to the behaviour of such experts.

A second way of expressing the worry that experts aren't possible stems from the alleged possibility that any subject might be a brain in a vat fed computer generated experiences straight into the brain. Let Fred be a putative expert about water and let Vat-Fred be an internal duplicate of Fred's brain placed in a vat. Suppose that when Fred has an experience of water, Vat-Fred has a computer-generated experience phenomenally identical to Fred's experience of water. More generally, suppose that if Fred has an experience E_i , then Vat-Fred has a phenomenally identical experience of vat- E_i and vice versa. Further suppose that Fred utters the words 'there is water in the lake'. Then, arguably, Vat-Fred is in an internal state which, if Vat-Fred were linked up to vocal chords, would cause his utterance of the very same words. If the inputs to Fred and Vat-Fred are phenomenally identical, then Fred cannot distinguish between them. Thus it seems that Fred cannot be an expert after all. Generalise this argument and you have an argument against the possibility of experts.

My response is to present the proponent of the argument with a trilemma. Suppose that we are experts about water and suppose further that, as a result of this, we can have water-thoughts regardless of the way the world is. The brain in a vat hypothesis dramatises the thought that, however much information about the world we obtain, we will not be able to decide between various alternative characterisations of its intrinsic nature. If we take our thought that there is H_2O (say) in the world as a claim about part of the intrinsic nature of the world, then the way to put this point is to say that we can have thoughts about the intrinsic nature of the world, and on

some accounts knowledge, but we will never be able to rule out the possibility that we are just receiving vat stimulations. If that is our situation, then our capacity to have such thoughts – and distinguish them from vat-possibilities in thought but not reality – displays the distinctive causal role of the thoughts in question. That's the first horn of the trilemma. The very capacity to have the thoughts that give rise to the sceptical problem indicates how these thoughts differ in (internal) causal contribution.

On the other hand, if we take our thought that there is H₂O as not making a claim about the intrinsic character of the world but just about the world of experience, then the possibility that we might be brains in vats does not establish that there may be two intentional properties with different causal powers (Putnam 1981, 12–17). Instead, ordinary subject and vat-subject will have the same intentional properties. For instance, Fred and Vat-Fred would have the same thought, a thought about H₂O. So there is no problem with their water-thoughts (say) having the same causal contribution to make. That's the second horn of the trilemma.

Only if we take our thought that there is H₂O as about whatever it is to which we stand in the right kind of relation – for instance, a certain kind of causal relation – would there be two distinct intentional properties, concerning water and vat-water, with the same causal powers. But now the claim that differences of intentional properties should correspond to differences in causal powers if they are to figure in scientifically advanced explanations comes under pressure. It remains true that the distinctions that we draw in terms of our ascriptions of intentional properties correspond to differences of causal powers. It is just that precisely which intentional properties we are ascribing is unclear since this is settled by the relations in which we stand to the environment (world or vat). This combination suggested by vat-cases is reproduced across the sciences. In characterising properties to figure in causal explanations, the sciences don't show any concern about distinguishing them from what might hold in vat cases. The causal profile of the properties we identify in our (let us suppose) non-vat world all have mirror properties with the same causal profile in the vat world. Nevertheless, this should not lead to us dropping explanations citing water (say). What matters is the identification of properties in terms of their ideally discoverable distinctive causal contribution. If this holds for the sciences in general, we shouldn't adopt another standard for intentional properties. That's the third horn of the trilemma.

The following scenario may appear to put pressure on this reply. Suppose that Fred's life is made up of a series of transplantations from body to vat. In such circumstances, Fred would be able to think about both water and vat-water. Yet his thoughts would not differ in their causal powers from Vat-Fred who undergoes a series of vat stimulations as if he is going through a life of transplantations. There are two points to make in response. If the issue is just one of comparing Fred and Vat-Fred regarding the causal powers of their thoughts, then the trilemma I sketched above applies just as much to this case. It's just that we are specifically concerned with the intrinsic character of the vat/non-vat contrast. On other hand, if the issue is a worry about whether Fred would be able to distinguish between his time in the real world – interacting with real water – and his time in the vat world – interacting with vat water – then whatever contextual reference fixing descriptions together with time of introduction that enables Fred to have thoughts about water rather than vat water, and vice versa, will be the basis for his distinction between water and vat-water. If he has failed to develop different concepts in this way because of his failure to distinguish vat world from real world, then he doesn't have different thoughts to display distinct causal powers.

So far I have only discussed the application of my proposal to intentional properties concerning kinds, specifically natural kinds though that was not essential to the discussion. It is possible to just stick at that and recognise that content-bearing states concerning individuals remain problematic. Indeed some are unconcerned about the causal explanatory role of one type of intentional property concerning individuals, those which determine indexical contents. They are presumably happy to bracket the differences of reference in, say, the use of 'I am sad' by different subjects, and focus on that which determines the common content expressed by these different uses (e. g. Owens 1987, 527–529). However, I think it would be helpful if I briefly outlined the way in which my strategy may be extended and any problems it would face.

There is no broad consensus about the proper way to characterise the indexical contents of content bearing states concerning individuals. This complicates my discussion. However, the moves I need to make are general ones which can be applied in different combinations to the variety of positions adopted. For my purposes, the main difference is between those who hold that the content of these states should be specified purely in terms of objects and properties

(e.g. Perry 1979, 1988) and those who propose that part of the specification of the content should be the mode of presentation of these objects and properties to the believer (Evans 1981a, b). That is, between those who hold that, for instance, the constituents of the content of Fred's belief *that that pigeon_{Mary} is pecking at his feet* should be represented as follows

(Mary, the property of pecking at Fred's feet)

and those who hold that it is to be represented by

(mp_{Mary} , mp the property of pecking at Fred's feet)

respectively (where ' mp_x ' is a mode of presentation of x). Those who characterise content in terms of objects and properties don't ignore the phenomena that led Frege to postulate modes of presentation (Frege 1892, 1918). They just deny that failures of substitution of coextensional terms in the ascriptions of propositional attitudes should be explained by supposing that these record differences in content. So when a subject believes that Cicero is a philosopher and does not believe that Tully is a philosopher (not knowing that Cicero = Tully), it is not that these two beliefs have different contents. There are other features that make the two beliefs different – for instance, different ways in which they are encoded or differences in their availability. Our ascription of different that-clauses to characterise the beliefs reflects this instead.

Since I am only interested in the causal powers of intentional properties, the extent to which differences in belief ascriptions capture differences in *content* is obviously quite important. But I need not settle the matter here so long as I can show that there would be a way of capturing the distinctive causal contributions of intentional properties whichever position was adopted. This forces me to explain both

(a) how differences in intentional properties due to them determining contents concerning different individuals correspond to differences in causal powers

and

(b) how differences between intentional properties due to one determining an indexical content and another a non-indexical content concerning the same individual (e.g. a content expressed by sentences containing proper names or definite descriptions at the relevant places) correspond to differences in causal powers.

Sticking with the example of Fred's belief *that that pigeon_{Mary} is pecking at his feet*, the first thing we need to do is capture the causal powers that this belief will have by being about a particular pigeon. Once more, the distinctive causal contribution will occur in the mind of the appropriate expert. The thought is that although *for many* one pigeon is like any other pigeon, there may be *some* who can distinguish Mary from all other pigeons: Mary experts. In them, a demonstrative belief that that pigeon_{Mary} is pecking at their feet will only result in a kick at Mary. Of course, qualitatively identical pigeons would be hard to distinguish by anybody's lights. An expert would have to appeal to their unique occupancy of a particular position in space and time or travel along a particular spatiotemporal path. Obviously, the merit of this suggestion rests upon whether it is even possible for a subject to discriminate occupancy of one space-time position from occupancy of any other. If it is not, the application of my proposal will be limited to intentional properties concerning kinds.

By contrast, the distinctive causal contribution of Fred's belief *that that pigeon_{Mary} is pecking at his feet*, compared with the belief *that Mary is pecking at his feet*, is brought out in the mind of the inexperienced. As has often been noted, it could come as news to Fred that the pigeon which is currently pecking around his feet, the pigeon which he thinks of as *that pigeon*, is Mary (e.g. Perry 1979, 1981). Before the news struck he would obviously behave differently if he had the belief that Mary was pecking at his feet and he wished to stop her from doing so, than he would if he just had the belief regarding *that pigeon_{Mary}*. If he wanted to chase off Mary, he would be frantically looking for the relevant bird amongst the cluster harassing him in St. Mark's Square. There would be no such search if he was just aiming to chase off *that pigeon*.

3. TWO VIRTUES OF CAUSAL EXPLANATION

In the previous Section I defended the claim that differences in intentional properties are differences in causal powers. It remains for me to explain how this blocks the *Argument from Precision*.

Suppose I provide a true causal explanation of the form
 e_1 having F causally explains e_2 having G

where F has its own distinctive causal powers. The valuable features of this explanation are the following.

- (i) It relates e_2 's causal history to all sorts of other causal histories involving events, states or objects with F, so showing e_2 's place in the causal network of the world.
- (ii) It unifies explanations of different phenomena. It records the fact that the occurrence of e_2 having G can be explained by citing a certain kind of event, state or object – one having the property F – that should be referred to in other explanations.

In addition, if F is discriminable by causal powers appropriate to a certain subject matter,

- (iii) It unifies the explanation of different phenomena *of a certain kind*.

For instance, suppose that a cause of Jo diving into a swimming pool is that he believes that it is full of water (not sulphuric acid). Jo is not an expert about water. He cannot distinguish between water and XYZ. A causal explanation of Jo's behaviour in terms of his belief that the swimming pool is full of water is not redundant because, in addition to citing a causally relevant property,

- (a) it relates that behaviour to a causal history that contains a belief – the belief that it is water – that has a unique slot to play in the causal network of the world;
- (b) it unifies the explanations of the various phenomena that arise from having both that particular belief and that kind of belief (individuated in terms of content);
- (c) it relates Jo's behaviour to other pieces of behaviour that might be the result of that particular belief and that kind of belief (individuated in terms of content) – e.g. (although I don't advise this) drinking the water of the swimming pool.

Obviously, this defence generalises. Given that features identified above are significant explanatory merits, it seems that there are virtues to explanations citing properties characterised in terms of causal powers which are not captured by focussing on precision alone.

Of course, similar points could be made if we were able to cite environment-independent intentional properties in an explanation. So it might be thought that the latter still win out since they have the virtue of precision too. In addition, it could be argued that there is a further benefit to citing environment-dependent intentional

properties. Classification by environment-dependent intentional properties misses generalisations that an alternative means of classification – appealing to environment-independent intentional properties or the non-contentful formal properties of mental states – would capture (e.g. see Stich 1983, 169–170; Loar 1988, 106–110). This last point threatens my claim that identification of environment-dependent intentional properties has the merit of unifying explanations of different phenomena. If this is a merit, it appears that we can go further in our attempts at unification.

Take two internally identical people who aren't expert about water. One is in a world with H₂O and the other in a world with XYZ. Suppose (for the sake of argument) that they share a common type of belief with an environment-independent intentional property. Let the belief be represented by N. One can think of the content of that belief encapsulating a state of knowledge that makes both subjects treat H₂O and XYZ indiscriminately, hard as that is to characterise. Then there will be a law of the form *ceteris paribus*, $\Box(x)(Nx \dots \supset Bx)$ – where ' \dots ' will be filled in with other mental states and background conditions, ' Bx ' stands for ' x behaves in way B', '*ceteris paribus*' captures the fact that psychological generalisations are not exceptionless, and ' \Box ' the fact that they have some sort of modality, most probably nomological. This law can be used to explain the behaviour of both subjects. In contrast, when we individuate by environment-dependent intentional properties, we would have to cite two laws: *ceteris paribus*, $\Box(x)(E_1x \dots \supset Bx)$; *ceteris paribus*, $\Box(x)(E_2x \dots \supset Bx)$ where ' E_1x ' and ' E_2x ' ascribe different environment-dependent beliefs to a subject x . The fact that we have two laws rather than one, it is suggested, means that we have missed a level of generalisation.

The objection has no force. First, arguing in favour of the legitimacy of explanations involving environment-dependent properties does not rule out also allowing that explanations involving environment independent properties are legitimate. Let Q be a feature that the two kinds of belief, E_1 and E_2 , share: as it may be, an environment-independent intentional property. Further suppose that when two subjects are not experts, this feature will make them respond the same way. Those who claim that intentional properties are environment-dependent can capture the generalisation, which their opponents held eluded them, by appeal to Q. The law they would provide is: *ceteris paribus* $\Box(x)(Qx \dots \supset Bx)$ where ' Qx ' stands for X has a belief with feature Q and one of the things that will be

implied by what is mentioned in place of ‘...’ is that x is not an expert.

The second point to make is that identifying beliefs by environment-*independent* intentional properties also loses generalisations. The expert and the inexperienced would be ascribed different environment independent intentional properties. So there will be two generalisations of the form: *ceteris paribus*, $\Box(x)(N_1x \dots \supset Bx)$ and *ceteris paribus*, $\Box(x)(N_2x \dots \supset Bx)$. Here ‘ N_1 ’ and ‘ N_2 ’ stand for beliefs with different environment-independent intentional properties. In contrast, there would be only one generalisation utilising environment-dependent beliefs, namely *ceteris paribus*, $\Box(x)(Ex \dots \supset Bx)$. This would capture what is common to expert and non-experts, their environment-dependent beliefs, and how it is related to their common behaviour. Indeed, as Fodor has pointed out, there may be sufficient individual differences between subjects that the *only* generalisation that abstracts away from these differences is one appealing to environment-dependent intentional properties (Fodor 1994, 49–54). In any event, things seem to be on a par regarding the ability of explanations citing only environment-dependent content-bearing states and those citing only environment-independent content-bearing states to capture or lose generalisations.

The third and final point to make is implicit in what I’ve argued already. Environment-dependent intentional properties have a different causal role to environment-independent properties. In which case, although each has the second explanatory virtue I have identified, they have it in virtue of the unification and characterisation of different patterns of causal explanation. So they are not in contest.

In which case, my conclusion stands. There are at least two causal explanatory virtues.

(I) The causal explanation of e_2 having G identifies a cause of e_2 having G , e_1 , by a property F such that (a) F has distinctive causal powers and (b) e_1 is a cause of e_2 having G in virtue of F .

(II) The causal explanation of e_2 having G identifies a cause of e_2 , e_1 , by a property F which is the most precise characterisation of what is necessary for e_1 to cause e_2 to have G .

Explanations which have either or both of the characteristics identified possess these virtues. No explanation which correctly and uniquely exemplifies either explanatory virtue is redundant. So the second premise of the Argument from Precision is without support. Explanations involving environment-dependent intentional properties are not redundant. Concern about the causal explanatory value of

citing environment-dependent intentional properties in causal explanations arose through a failure to recognise any other virtue than precision. There is no argument from explanatory redundancy to the rejection of the Externalist approach to intentional properties.³

NOTES

¹ I'm grateful to Tom Stoneham for making me think about the second possibility.

² Even if it were, the causal powers of a supervening property may outstrip those of its realisation (see Noordhof 1997, 1999a).

³ Many thanks to the audiences in the Philosophy Departments of King's College London and York and to the members of Alexander Bird, Johannes Persson and Helen Beebe's Metaphysics and Philosophy of Science group where I presented this paper. Their questions were invaluable. I am also very grateful for the comments of Michael Clark, Eros Corazza, Bob Kirk, Greg McCulloch, Hugh Mellor, and Alex Miller.

REFERENCES

- Burge, T.: 1979, 'Individualism and the Mental', in P. A. French, T. E. Uehling, and H. K. Wettstein (eds.), *Midwest Studies in Philosophy* Vol. IV, pp. 73–121.
- Burge, T.: 1986, 'Individualism and Psychology', *The Philosophical Review* **95**, 3–45.
- Burge, T.: 1989, 'Individuation and Causation in Psychology', *Pacific Philosophical Quarterly* **70**, 303–322.
- Burge, T.: 1995, 'Reply: Intentional Properties and Causation', in C. Macdonald and G. Macdonald (eds.), *Philosophy of Psychology*, Basil Blackwell, Oxford, pp. 226–235.
- Dretske, F.: 1996, 'How Reasons Explain Behaviour: Reply to Melnyk and Noordhof', *Mind and Language* **11**, 223–229.
- Egan, F.: 1991, 'Must Psychology Be Individualistic?' *The Philosophical Review* **100**, 179–203.
- Egan, F.: 1992, 'Individualism, Computation, and Perceptual Content', *Mind* **101**, 443–549.
- Evans, G.: 1981a, 'Understanding Demonstratives', in H. Parrett and J. Bouveresse (eds.), *Meaning and Understanding*, (reprinted in his: 1985, *Collected Papers*, Oxford University Press, Oxford, pp. 291–321).
- Evans, G.: 1981b, *Varieties of Reference*, Oxford University Press, Oxford.
- Fodor, J.: 1991, 'A Modal Argument for Narrow Content', *Journal of Philosophy* **88**, 5–26.
- Fodor, J.: 1994, *The Elm and the Expert*, The MIT Press, Cambridge, Massachusetts.
- Frege, G.: 1892, 'On Sense and Meaning', in Brian McGuinness (ed.): 1984, *Collected Papers on Mathematics, Logic and Philosophy*, Basil Blackwell, Oxford, pp. 157–177.
- Frege, G.: 1918, 'The Thought: A Logical Inquiry', *Mind* **65** (1956), 289–311 and in P. F. Strawson: 1967, ed., *Philosophical Logic*, Oxford University Press, Oxford, pp. 17–38.

- Hornsby, J.: 1986, 'Physicalist Thinking and Conceptions of Behaviour', in P. Pettit and J. McDowell (eds.), *Subject, Thought and Context*, Oxford University Press, Oxford, and in her 1997, *Simple Mindedness*, Harvard University Press, Cambridge, Massachusetts, pp. 111–128 [page references in text to latter].
- Loar, B.: 1988, 'Social Content and Psychological Content', in R. Grimm and D. Merrill (eds.), *Contents of Thought*, University of Arizona Press, Tucson, pp. 99–110.
- McGinn, C.: 1989, *Mental Content*, Basil Blackwell, Oxford.
- Noordhof, P.: 1997, 'Making the Change: the Functionalists way', *The British Journal for the Philosophy of Science* **48**, 233–250.
- Noordhof, P.: 1999a, 'Micro-based Properties and the Supervenience Argument: a response to Kim', *Proceedings of the Aristotelian Society* **99**, 109–114.
- Noordhof, P.: 1999b, 'Causation by Content', *Mind and Language* **14**, 291–320.
- Owens, J.: 1987, 'In Defence of a Different Doppelganger', *The Philosophical Review* **96**, 521–554.
- Peacocke, C.: 1981, 'Demonstrative Thought and Psychological Explanation', *Synthese* **49**, 187–217.
- Peacocke, C.: 1993, 'Externalist Explanation', *Proceedings of the Aristotelian Society* **93**, 203–230.
- Perry, J.: 1979, 'The Essential Indexical', *Nous*, **13**, 3–21, and in his 1993, *The Essential Indexical*, Oxford University Press, Oxford, pp. 33–50.
- Perry, J.: 1988, 'Cognitive Significance and New Theories of Reference', *Nous* **22**, 1–18, and in his: 1993, *The Essential Indexical*, Oxford University Press, Oxford, pp. 227–247 [page references in text to latter].
- Putnam, H.: 1981, *Reason, Truth and History*, Cambridge University Press, Cambridge.
- Segal, G. and E. Sober: 1991, 'The Causal Efficacy of Content', *Philosophical Studies* **63**, 1–30.
- Shoemaker, S.: 1980, 'Causality and Properties', in Peter Van Inwagen (ed.), *Time and Change*, D. Reidel, Dordrecht, pp. 109–135; reprinted in S. Shoemaker: 1984, *Identity, Cause and Mind*, Cambridge University Press, Cambridge, pp. 206–233 [page references in text to latter].
- Stich, S.: 1983, *From Folk Psychology to Cognitive Science*, The MIT Press, Cambridge Mass.
- Wilson, R.: 1992, 'Individualism, Causal Powers and Explanation', *Philosophical Studies* **68**, 103–139.
- Wilson, R.: 1995, *Cartesian Psychology and Physical Minds*, Cambridge University Press, Cambridge.

Department of Philosophy
University of Nottingham
University Park
Nottingham NG7 2RD
UK
E-mail: paul.noordhof@nottingham.ac.uk