What We Do to Each Other
Interventionism, Folk Psychology, and Special Causal Concepts

Abstract

It is part of our ordinary understanding of one another that our mental lives are richly causally structured. But reconciling the causal commitments of our folk psychology with one or another philosophical account of causation typically leads to a problem of 'mental causation'. An 'interventionist' approach to causality, as pioneered by James Woodward and others, promises to do better. On an interventionist conception, causal relationships are associations that are preserved in hypothetical circumstances when the putative cause is acted on by an exogenous factor. In this paper I argue that this promise remains unfulfilled. The basic reason is that our lives are too closely causally intertwined, and our interactions too much dependent on a shared background of mutual intelligibility, for normal actions towards one another qualify as interventions in the relevant, quasi-technical sense. And this makes it hard to see how our causal understanding of one another could involve an appreciation for information about what would happen under hypothetical interventions.

This is not just a technical difficulty for the interventionist approach; rather, seeing why the paradigm is inapplicable here sheds light on the intuitive sense that interpersonal folk-psychological understanding is not detached or disengaged, but rather grounded in reciprocity. One response would be to hold that interpersonal understanding, insofar as it is so grounded in mutuality, is not causal. In the latter part of this paper I sketch an alternative, pluralist approach, inspired by G. E. M. Anscombe's discussion of special causal concepts in her essay 'Causality and Determination'. I close with some remarks about how this proposal allows us to reconceive the relation of folk psychology to the metaphysics of mind.

It is part of our ordinary understanding of ourselves that our psychological lives are causally structured. Mainstream philosophy of mind has tended to focus almost exclusively on the causal role of just two aspects of this everyday understanding, namely belief and desire (and perhaps also—on sufferance—intention.) But a quick glance at some familiar and recognisable psychological occurrences shows a much more varied inventory of the causal goings-on of mental life: lashing out in anger, being irritated by the tone of someone's voice, having a childhood memory jogged on revisiting a familiar place, getting distracted from one's work by pangs of hunger, suppressing an urge to tell someone what you really think of them, having one's spirits revived by a conversation with an old friend—and so on. Much of our vocabulary for describing psychological life also makes liberal use of mechanical and hydraulic language, hinting at a really quite rich and textured conception the causal dynamics of the mind: we speak easily of being overwhelmed by a flood of pent-up emotion, the mental strain involved in suppressing one's negative feelings about a situation, labouring under the weight of a personal loss, or one's resolve cracking under mounting pressure. So in our ordinary conception of ourselves we seem to be deeply committed to the causal character of a whole armoury of psychological notions.

The questions I am concerned with are: What is involved in ordinary folk understanding of psychological causality? What does the causal content of our folk psychological explanations amount to? And what would it take for these explanations to amount to genuine knowledge of causality in the mind?

Somewhat surprisingly, these questions have not received a great deal of attention in the vast literature on folk psychology and mindreading.[1] Questions about the acquisition and possession of folk-psychological concepts—like those of belief and desire—have been much discussed, and it is typically assumed that someone who does possess these concepts will be able to deploy them in giving causal explanations of psychological phenomena. But it is rarely questioned what kind of understanding of causality is implicated in this conceptual capacity.

Perhaps a reason for this gap is an assumption that the concept of cause is

[1]For just some significant contributions to this literature, see Apperly 2010; Carruthers and Smith 1996; Goldman 2008; Nichols and Stich 2003.

a single, general-purpose concept. On this assumption, causal psychological understanding is just a matter of combining psychological concepts with a domain-general notion of causal influence. For example, many discussions of 'mental causation' in the philosophy of mind have supposed that causally explaining a phenomenon is a matter of deducing its occurrence from general laws, and that the same goes whether the phenomenon in question is physical or psychological.

On the other hand, there is by now a substantial body of work in developmental and cognitive psychology on human causal learning and cognition, in which it is by no means taken for granted that whatever the dominant philosophical account of causation is—for instance, a law-based account—corresponds to anything in how humans represent and reason about causal relations.[2] For example, an important question is to what extent human causal cognition relies on domain-specific 'substantive assumptions' about what can cause what, such as no-action-at-a-distance locality, or whether causal connections are rather inferred purely from observed contingencies in accordance with a few schematic rules. These questions might be glossed as probing what kind of concept of cause humans actually deploy in learning about the world around them: a mechanistic, transfer-based concept, or a more 'Humean' concept on which causes are just a special kind of regularity.[3]

Thus, when it comes to our understanding of causality in the mind, we should similarly view it as an open question just what the relevant understanding of causality amounts to: what, if any, constraints it is subject to, what connections it bears to other, neighbouring concepts, what kinds of information its claims are based on. These questions are pressing because our experience of the psychological world is in many ways quite different from that of the physical. For instance, as Gopnik and Meltzoff point out, we typically influence one another psychologically by means of communication; hence, unlike in the physical case, "'action at a distance" is the

[2]For some collections drawing together empirical and philosophical perspectives on causal cognition, see Gopnik and Schulz 2007; Hoerl, McCormack and Beck 2011; McCormack, Hoerl and Butterfill 2011; Sperber, D. Premack and A. J. Premack 1995; Waldmann 2017. Woodward 2007, 2021 represent an important attempt to draw together questions about causal cognition and the metaphysics of causation, broadly in the context of the 'interventionist' framework that I go on to discuss.

[3]See e.g. Hoerl 2011 for discussion of specifically this issue.

rule rather than the exception in psychological causality.' (Gopnik and Meltzoff 1997, p. 141)

Indeed our understanding of psychological life seems on the face of it so different from our grasp of the physical world that it is reasonable to ask whether the concept of cause has any place in it at all. There are two prominent negative answers to this question. One is Daniel Dennett's 'intentional stance' (Daniel C. Dennett 1981, 1991; Daniel Clement Dennett 1981). On Dennett's view, folk psychology predicts and explains behaviour by parsing it into goals and rational, informationally constrained means. It thus in a certain limited sense treats behaviour as the product of relevant beliefs and desires, but without being committed to such beliefs and desires as genuine psychological causes. A different negative answer, associated principally with certain followers of Wittgenstein and with the *Verstehen* tradition in the philosophy of social science, says that understanding a human action is a matter of interpreting it, by situating it in a web of socially constituted meanings, or by characterising it as a move in a social game, and that this is a distinct project from explaining an action causally in terms of the chain of psychological happenings that led up to it.[4] On both of these approaches, there may be certain causal questions about the deep causes of someone's actions—for instance, whether someone acted out of love or spite—that come out as simply indeterminate, corresponding to no real causal distinction.

The visibility of these options helps sharpen our question. Just what is it that is added by claiming that, say, love rather than spite was the real cause of someone's action? What, if anything, do claims of this kind have in common with ordinary claims of physical causality, like that the brick smashed the window? And, finally, what would it take for the claims of psychological causality to be vindicated?

An 'interventionist' approach to causation, as pioneered recently by James Woodward (Woodward 2003) and others (e.g. Halpern and Pearl 2005; Hitchcock 2001; Spirtes et al. 2000), offers a promising basis for a principled set of answers to these questions. In particular, in a series of papers, John Campbell has defended precisely an interventionist approach to psychological causation (Campbell 2006,

[4]The classic texts here are Davidson's 'little red books': Kenny 1963; Melden 1961; Winch 1958. Some influential theorists of interpretation in the social sciences are Geertz 1973; Taylor 1971.

2008, 2010).[5] The basic idea of this approach is that causal relationships are regularities that continue to hold when the putative cause is manipulated by an exogenous source, where the paradigm of an appropriately exogenous manipulation is a controlled or randomised scientific experiment.

This approach has a number of attractions. It provides a pragmatic answer to the question what we gain by representing one another's psychological lives in causal terms: on an interventionist approach, tracking causes is functional in a particularly straightforward way because it allows us to identify those relationships that are relevant to predicting the results of our own actions and manipulations, and—so it might be said—this is no less the case for our interactions with other people than with inanimate nature. Secondly, interventionism meshes nicely with a plausible epistemics of causal understanding, whereby we learn about what causes what by noting the results of our actions on objects around us—including our fellow humans—and thus promises to makes it unmysterious how we could could come to acquire genuine causal knowledge of the mind. More generally, interventionism is currently the best-developed version of a 'difference-making' approach, on which causality is understood in terms of patterns of dependencies and contingencies, rather than in terms of physical mechanisms. This seems like a welcome feature, since many of the hallmarks of physical mechanisms, such as spatiotemporal contact and the transfer of energy and momentum, do not apply to the psychological case; indeed the whole idea of a psychological 'mechanism' in general is notoriously obscure.

In this paper I critically assess the extent to which an interventionist approach does indeed capture anything of our everyday causal understanding of one another. I argue that it can do so only to a limited extent. What interventionism gets right is the attractive thought that understanding one another causally is intimately connected with being able to intervene in one another's psychological lives, paradigmatically via communication. The problem is that our interactions with one another are very far from the ideal of an exogenous manipulation. For this reason, try-

[5]Campbell 2020 is more equivocal on interventionism, and in the position he develops there is in many ways similar to the positive alternative sketched in §6—although I will not undertake an explicit comparison here.

ing to fit folk psychological understanding into an interventionist mould leads to spurious sceptical problems, insofar as our everyday patterns of interactions with one another do not give us good reason to think that the relevant interventionist conditions of causal influence hold. Accordingly, I suggest that we need to adopt a more pluralistic perspective on causal understanding, on which folk psychology consists of an array of *sui generis* 'special' or 'thick' causal concepts.

After briefly characterising interventionism and its advantages (§1), I motivate, on general grounds, the claim that ordinary interpersonal actions are not typically interventions by one person in another's mental life, in the relevant technical sense (§2). The following sections unpack this claim in more detail, and spell out exactly what the problem is, with reference to examples (§§3–4). After a recap and appraisal (§5), I sketch an alternative approach, on which folk psychological concepts are thick causal concepts (§6). Finally I make some remarks on the ramifications of this proposal for the metaphysics of psychological causality (§7).

1 The attractions of interventionism

The broad idea behind an interventionist approach to causality is that what distinguishes causation from mere correlation is that causal relationships are just those correlations which continue to hold when the cause is acted on by an exogenous source. This is meant to codify a pattern of inference that is ubiquitous in science and in informal empirical reasoning. Suppose we have noticed the regular co-occurrence of two phenomena, A and B, and want to know whether A causes B, or B causes A, or they are both effects of a common cause; or whether the association is merely accidental. One thing we might do is contrive to bring A about, not by its normal route, but through our own efforts, and see if B still happens. If it does, then we can typically infer that A genuinely causes B, rather than there being some other explanation of the regularity. The interventionist approach makes this procedure definitive of what causation is: genuinely causal relationships are just those regularities that are *robust to intervention on the cause*.

The question is then what it takes for an intervention to be appropriately 'ex-

ternal'. The principal innovation of the interventionist approach is to define the notion of an intervention, without reference to human agency, as an event with a certain kind of causal structure.[6] The idea is that, when we make causal inferences from our interventions, what matters is not that the interventions are our actions as such, but that they are exogenous in the right kind of way. First off, we have to actually succeed in bringing about A, rather than A just coming about in the usual way, or not at all. Secondly, whatever we do to bring about A must not have some additional effect, C, unconnected with A, which also affects B. Finally, our attempt to bring about A must not be correlated with some other background feature, Z, that affects B independently of A. The paradigm of an intervention that satisfies these conditions is a randomised independent variable in a controlled scientific experiment. In randomising and controlling, we are aiming to reproduce a natural phenomenon in order to observe just its effects, striving to eliminate all the noisy and confounding additional features that are present when it occurs in the wild.

This can be made more precise by introducing the notion of a *causal structure*. The causal structure of a system is the set of causal relationships between its various elements. The kind of causal relationship meant here is not just a one-off instance of one particular event causing another, but nor is it a wholly general type-level claim, like that smoking causes cancer. Rather, it refers to the systematic dependencies that hold between aspects of some specific system. A model for this grade of causal involvement might be the sense in which we would say that one part of a mechanical contraption like a clock or a car engine moves another part, not on any particular occasion, but as a matter of how the parts of the machine are structurally connected up. But this model should not mislead—the distinct elements of a causal structure need not be spatially separated component parts, and the causal relations between them need not be mechanical ones.

Causal structure can be helpfully represented by means of *directed acyclical graphs* (DAGs). Formally, a DAG is an ordered pair $\{V, R\}$ consisting of a set of

[6] As has been often emphasised, this means the interventionist definition of cause is not reductive. However, it is not trivial or viciously circular, because, although the notion of cause features in the explicit definition of 'A causes B', the condition that A causes B does not. Cf. Woodward 2003, pp. 104–107.

variables V and irreflexive, acyclical relation R over V. The variables comprised by V correspond to determinable aspects of the represented system, and their values correspond to the specific states or properties which that aspect of the system might instantiate. For example, there could be a binary variable representing the position of a switch, or a continuous variable representing the pressure in a combustion chamber. The relation R represents causal-structural relations between these determinable elements.

Here is a DAG for a simple kind of causal structure, one in which X and Y are joint effects of a common cause Z. Here $V = \{X, Y, Z\}$ and $R = \{\langle \text{Z,X} \rangle, \langle \text{Z,Y} \rangle\}$. A stock example of this causal structure would be one in which X represents the position of a barometer needle, Z represents local atmospheric pressure, and Y represents the occurrence of rain.

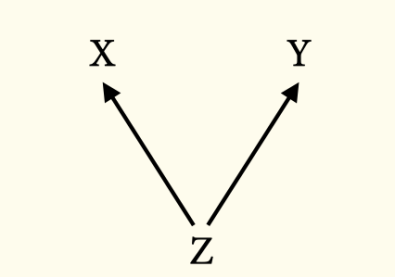

Figure 1: Common cause structure

An intervention is then defined in terms of the causal-structural notion of an *intervention variable*. An intervention variable for X with respect to Y is a variable I such that:

I1 I causes X, i.e. there is a chain of arrows or *directed path* leading from I into X.

I2 I acts as a 'switch' for X, i.e. when I=i, X always takes some specific value $x*$, so that the value of X is (statistically) independent of the usual causes of X.

I3 I only affects Y via its influence on X, i.e. there is no directed path from I into Y that does not go through X.

I4 I is not correlated with any additional variable V that is on a directed path to Y that does not go through X.[7]

The effect of an intervention can be represented graphically by drawing an arrow from I into X and removing all other arrows into X. Fig. 2 shows the same causal structure under intervention on X, and fig. 3 shows an improper 'intervention' in which I* is correlated with some independent cause of Y, violating condition I4 (the undirected edge represents a correlation, without conveying any information about causal priority.)

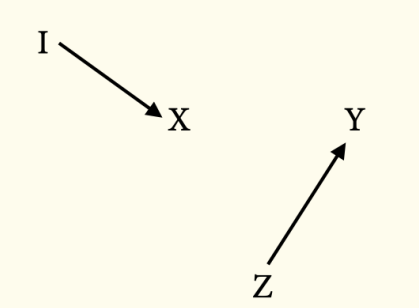

Figure 2: Common cause structure, intervention on X

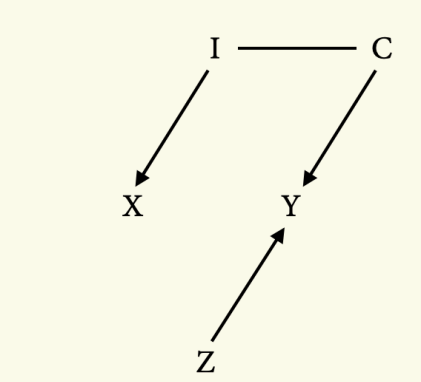

Figure 3: 'Confounded' intervention on X (I4 not satisfied)

It is easy to see why an event like I is revealing of structural-causal relations between X, Y, and Z, whereas a defective intervention as in fig. 3 will not do. Normally, X and Y are correlated, but this is due to the common cause Z rather than any causal influence between them. So, if we intervene to set X externally, overriding its 'normal' cause Z, we should not observe the usual concomitant change in Y—the 'spurious' correlation between them is broken. On the other hand, if, as in fig. 3, our intervention is 'confounded' by a further cause of Y that is also correlated with the intervention, some correlation may remain between X and Y—we not will not have succeeded in prising X sufficiently cleanly away from the web of causal relations in which X and Y normally occur. This is exactly what happens

[7] This is essentially the definition from (Woodward 2003, p. 98). For an alternative characterisation see Pearl 2009, p. 108.

in a badly designed study when the experimenter's actions are covertly influenced by some feature of the situation other than the one whose effects are being studied, like the participants' social class, ethnicity, or gender presentation, which may be correlated with other factors that are relevant the outcome.

To avoid confusion, in what follows I will refer to an event that satisfies conditions I1–4 as an Intervention. This makes clear that it is an open question whether a given action or event that we might informally call an 'intervention' is an Intervention in this specific sense.

The primary interventionist notion of cause makes it a structural, or variable-level, phenomenon: a special kind of dependency relation between aspects of a system. However, it is also possible to define, in interventionist terms, a notion of particular, or 'actual', causation between specific occurrences. This is essentially a counterfactual analysis: an event *c* is a cause of an event *e* just in case, had some Intervention been carried out to alter or prevent *c*, *e* would have occurred differently or not at all. The more precise content of this interventionist counterfactual is then spelled out in variable-level, causal-structural terms. A somewhat simplified definition runs that an event *c* is an actual cause of an event *e* just in case:

AC1 *e* is an event $X = x$ of some variable X taking the value *x*, and *c* is an event $Y = y$

AC2 There is some possible event $I = i$, in which an Intervention variable I for X with respect to Y takes the value *i* and thereby causes X to take the value $x\prime$.

AC3 If $I = i$ were to occur, but all other causes of Y remain at their actual values, then Y would take some different value $y\prime$.[8]

As noted above, interventionism boasts a number of advantages when it comes to capturing the causal claims of folk psychological explanation. In common with other difference-making theories, interventionism is a quite liberal and permissive

[8] A full-dress definition of actual cause needs some further refinements to cover various kinds of overdetermination and pre-emption, e.g. Woodward 2003, p. 84; also Hall 2000; Hitchcock 2001.

theory of causal explanation that promises to secure the causal relevance of high-level properties and events. It eschews many of the more demanding *a priori* requirements of many philosophical and scientific theories of causation such as laws, conservation of quantities like energy or momentum, and so on. More generally, it does not require in any substantive sense that there be any specific mechanism (physical or otherwise) connecting cause and effect. The only notion of a 'mechanism' is just that of a route of causal influence, characterised in interventionist terms, as a factor that can be separately tweaked or manipulated to modulate a certain outcome. Given the obscurity surrounding the notion of a causal mechanism as applied to the psychological realm, this seems like a very welcome feature.

Beyond this, interventionism claims a considerable advantage over alternative difference-making theories when it comes to the question of causal understanding. These advantages come out, for instance, when comparing interventionism to other counterfactual theories, such as David Lewis's influential version (Lewis 1987). Counterfactual theories face the general challenge of securing the correct interpretation of the relevant counterfactuals. In particular, there needs to be some principled basis for eschewing 'backtracking' reasoning like 'If the barometer needle had been in a different position, the air pressure would have been different; so it would have probably rained.' This kind of conditional reasoning may be intuitively acceptable, at least in certain contexts; but it will not do for an analysis of causation, which requires us to 'hold fixed' all factors prior to the occurrence of the cause. In Lewis's account, the non-backtracking interpretation is secured by a complicated metric of similarity across possible worlds. Even if Lewis's metric yields an extensionally correct analysis, though, the problem when it comes to causal understanding is that it is hard to explain how we ever came to be interested in this cross-world similarity relation rather than some other one; or how we should go about assessing whether a given causal counterfactual holds.

By contrast, interventionism is able to relate causal counterfactuals systematically to information about certain special associations between *actual-world* events, namely between Interventions and their outcomes. It thus has a convincing story about why causal counterfactuals are of interest in the first place, because they

tell us what to expect when we perform actions that are Interventions—as many of our actions are. Moreover, it promises to demystify the interpretation of the causal counterfactuals: there is no need to stipulate and explicate a special 'non-backtracking' interpretation of the counterfactual, because if the event mentioned in the antecedent is an Intervention, then, by I4, its occurrence is independent of anything else that might affect the outcome event, and so any backtracking is irrelevant to the evaluation of the consequent.

Of course, the official notion of an Intervention is quite technical, and it is not plausible to suppose that many people have this in mind when they engage in causal reasoning. The point is that interventionism offers a regimented and precise characterisation of a 'real pattern' that we are sensitive to, albeit not in those very terms, in our causal and practical reasoning. Here is Woodward on how human causal cognition might follow a basically interventionist mould:

> ...human beings (and perhaps some animals) have (i) a default tendency to behave or reason as though they take their own voluntary actions to have the characteristics of interventions...and, associated with this, (ii) a strong tendency to take changes that temporally follow those actions with a short delay as caused by them. If (iii) the default tendency in (i) is often correct (or if we are fairly good at recognizing when it is likely to be correct), then, on interventionist principles, (iv) the tendency in (ii) will also often be correct. (Woodward 2021, pp. 208–209)

We can see how someone satisfying these conditions could thereby manifest a basically interventionist understanding of causal claims, and the associated counterfactuals, without being able explicitly to formulate the technical notion of an Intervention, but rather in the patterns of planning and inference that they engage in as agents interacting with their environment.[9] If an interventionist account of causality is correct, then these inferences will be warranted, and can be counted as a source of genuine causal knowledge rather than merely, say, a useful planning heuristic. Interventionism therefore offers a principled account, first, of the patterns of use that characterise our everyday grasp of causality and, secondly, how those patterns of use can be genuinely knowledge-generating.

[9]There is a tricky issue here about whether we should count an agent whose action and inference patterns meet these conditions as possessing the *concept* of cause, that is, as representing causal information in a conceptual format. For discussion of this point, see Hoerl 2011.

These armchair considerations are supported by the developmental evidence that interventionist-style learning by doing plays a key role in young children's learning about causation. A body of experimental work carried out by Alison Gopnik and colleagues shows children to be capable of making causal inferences on the basis of theirs and others actions, which typically at least approximate to Interventions (Gopnik, Glymour et al. 2004; Gopnik and Kushnir 2003; Gopnik, Kushnir and Schulz 2007). This work so far has predominantly concerned children's learning about the physical world; but it is not a great stretch to suppose that much the same learning processes are at work when children observe the effects of their expressions, gestures and utterances on their caregivers, constructing causal psychological theories to explain their communicative successes and failures.

Here is a simple example which might be seen to exhibit a basically interventionist notion of psychological causality, and which I will return to frequently in the ensuing discussion:

Directions to the station As you are walking around the city centre, someone runs up to you and breathlessly asks the way to the train station. You point down the road to the west and they shoot off in that direction.

A very natural causal interpretation of this scene is that you have given the person the belief that train station is to the west, and that this belief has in turn caused them to run that way. On an interventionist gloss, the causal connection between the belief and the outcome behaviour is articulated in terms of the idea that a different communicative action would have produced a different belief, which would have been accompanied by a correspondingly different outcome behaviour: had you pointed in some other direction, they would have acquired a different belief, and run off in a different direction. This is enough for genuine causal knowledge, because it is all there is to the causal connection: there is no need for anything deeper, like knowledge of the neurophysiological laws and mechanisms that led up to the person's running off in that direction, in order for your causal beliefs to be vindicated.

In summary, interventionism appears far better-placed than many other leading theories of causation and causal explanation to capture, and to vindicate, our ordinary understanding of causality in the psychological realm. Nevertheless, despite these appearances, it still faces significant challenges. This is the topic of the next few sections.

2 The causal structure of interpersonal action

The basic problem with interventionism, as an account of the causal content of folk psychology, is this: the canonical way in which we intervene on one another's psychological states is via communication; but our communicative actions generally fail to qualify as Interventions (in the relevant technical sense.) For this reason, there is no straightforward relationship between the ways in which we ordinarily understand ourselves to be capable of influencing one another, and information about how people's behaviour would be affected by hypothetical Interventions. And this puts in doubt the ability of interventionism to capture or vindicate the causal content of folk psychology. This and the following two sections will develop this claim in detail.

I do not think it is possible to give a knock-down proof that a communicative action cannot be an Intervention. But there are some general and widespread features of the mind, and the ways we ordinarily influence each other, that stand in marked contrast to the ideal of an experimental manipulation.

The first point is that when we influence other people communicatively, we do not typically manage to fix a given psychological variable—a particular belief, say—in a way that renders its normal causes irrelevant. Rather, we normally influence people by giving them reasons for a new belief or course of action. In doing so, we are not overriding a person's own endogenous processes of deliberation and belief fixation, but rather competing, or collaborating, with them. We impinge on the whole causal psychological web, in which practically anything can be causally relevant to anything else, rather than surgically isolating a single feature. This point is observed by Campbell, who notes, 'It does not happen very often, if it happens

at all, that a person's rational autonomy is suspended and some alien force seizes control over whether that person has a particular intention.' (Campbell 2006, p. 61)

The second important point is that interpersonal actions are typically not at all random or arbitrary, but rather arise, semi-endogenously, from a background of mutual knowledge and interaction. Imagine walking around a small market town with a casual acquaintance, vaguely chatting, commenting on buildings and people you pass, and so on. This is an activity we readily think of as a good way of getting to know someone better. Intuitively, it seems completely wrong to think of your casual remarks as mini-experiments designed to probe the causal dynamics of the other person's mind. And the reason for this, I suggest, is to do with your motivation in saying what you do: rather than being an act of external interference, each of your contributions arises naturally out of the evolving context of your mutual acquaintance.

This one example does not prove that there are no interpersonal actions which are Interventions. But there are pertinent features that hold for communicative actions across the board. Making a difference to someone's psychology by communicating with them requires a significant amount of stage-setting and pre-alignment in order to come off. If the participants are badly out of sync, there is a danger that any attempted communication will fail to produce the intended psychological effect, and instead result in disruption or confusion. At a minimum, the participants must have a functioning understanding of the relevant communicative conventions, typically including command of a common language, in order to understand each other. They also need enough of a common understanding of the local setting to interpret context-dependent signals, including lexical items such as pronouns. This may mean directing perceptual attention to the same environmental features, or having a common conception of a remote subject-matter. Beyond this, there needs to be a more general widespread agreement about what the world is like and what kinds of things make sense, or else they may surprise each other too much for the conversation to function. Finally, there needs to be shared adherence to conventions of conversational interaction, such as turn-taking norms, in order to ensure the flow of conversation does not break down. This might mean not only

conscious observance of explicit norms, but also finer and lower-level perceptually guided motor skills, such as appropriate control of one's bodily position, hand gestures, and facial expression, maintenance of personal space, regulation of eye contact, and so on. Human communication—even in more adversarial settings—is a meticulously fine-tuned dance that would seem to any alien observer like a miracle of co-ordination.

From a purely causal point of view, then ordinary communication looks less like a controlled external manipulation and more like a burst of activity in a densely interconnected feedback system. The twin problem this creates is then that the same background features of the situation, or correlated background features like individual beliefs with the same content, play a causal role in determining both what one person says and how the other person responds. The next few sections spell this out in detail.

3 Switches and nudges

The first point was that, when you interact with someone, you cannot grab hold of some aspect of their psychological state and fix it in a way that bypasses normal deliberation. So condition I2, that an Intervention should act as a switch for the target variable, is not satisfied by ordinary communicative interventions. An event of this kind, depicted in fig. 4, is at best a 'soft' Intervention: rather than a switch that grabs control of a variable, more of a 'nudge' that exerts an influence on the variable, but without totally suspending or overriding the influence of its other normal causes.

Go back to Directions to the station. You cannot just magically see to it that they get the belief the station is to the west. You only bring about the belief via other, intervening psychological changes, such as the perception of your pointing in that direction. And this in turn causes a new belief only in concert with other background beliefs, such as the presumption that you understood the question, and that if you understood the question correctly and are pointing to the west, the station is probably to the west. The role played by these presumptions also plaus-

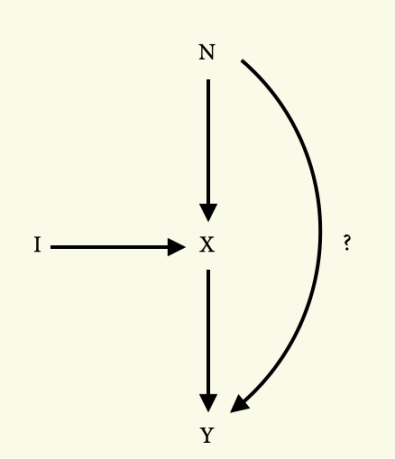

Figure 4: 'Soft' intervention on X

ibly limits your ability to produce, by means of your intervention, certain beliefs about the subject matter in question: for example, if you pointed straight at the sky, you would be unlikely to cause the corresponding belief that the station is in that direction.

The problem is then this: given that you have not totally suspended the normal causes of, for instance, someone's belief about where the train station is, it is possible that those normal causes might have brought about the outcome anyway, quite independently of your action. Although in the above situation you might be confident enough that this is not the case, it is not clear, on the official interventionist account of causation, why this conviction of the causal efficacy of your intervention should be anything more than a kneejerk prejudice—like getting a call from an old friend you happen to be thinking about, and superstitiously thinking you somehow 'made' the friend call. After all, it is not as if the above situation would pass muster as a well-designed psychology experiment.

3.1 Campbell's amendment

Campbell accordingly suggests we need a more nuanced version of interventionism, one that relates causal claims to information about 'soft' Interventions—nudge-like events which satisfy I1, I3 and I4, but not I2. The basic idea is that, rather than requiring that the Intervention make the normal causes of the target

variable irrelevant, instead we take into account the actual values of those normal causes when we look at the difference the Intervention made to the outcome variable. We can do this by comparing the value of the outcome variable Y when we intervene on X with the value of Y absent our intervention, given the actual state of the normal causes of X. Campbell explains: 'We are not any longer considering whether the value of Y is independent of the value of X, when the value of X is set by surgical intervention. We are, rather, considering whether Y is independent of the intervention variable I given the usual causes of X.' (Campbell 2006, p. 65)[10]

Here is an example. Suppose you want to know how much your plants' growth is affected by the moisture of the soil. You might set up a variable watering regime, and note down how the rate of growth of each plant changes varies with the amount of water you give it. But if the plants are not in a controlled environment, there may be other causes of the level of moisture other than your watering, such as precipitation and ambient temperature—and these factors (or their correlates, like sunlight) may have an independent effect on growth. This is the situation depicted in fig. 4. If so, then by watering the plant you have not rendered soil moisture independent of factors like the weather, so any association between them cannot be assumed to reflect just the influence of soil moisture on growth. Campbell's suggestion is that the thing to look at is therefore not whether there is an association between soil moisture and growth when you water each plant, but rather the difference between the observed rate of growth when you watered the plant, as compared with the growth you would expect if you had not watered it but the other endogenous causes of soil moisture, like the weather conditions, all remained the same.

Note that for this amended version of interventionism to be practicable as a way of detecting causal relations, it has to be possible for I to vary independently

[10](Kaiserman 2020) suggests a less radical amendment: we stick with the standard interventionist definition of cause in terms just of correlations under intervention, but amend I2 to allow cases where I affects X 'indirectly', via its normal causes—ruling out only those cases, like in fig. 4, where the normal causes affect X independently of I. Kaiserman points out that, in cases of 'indirect' intervention, any covert affect of the intermediate N on Y that does not go via X would be in violation of I2, and so already taken care of. While this solution works in principle, the problem is that we have every reason to think that cases like fig. 4 are widespread in interpersonal life: when we affect someone's conduct via communication, our interventions are always susceptible to being undermined, overridden or amplified by the other person's endogenous cognitive processes.

of the normal endogenous causes of X, so that we can consider a range of different hypothetical Interventions given the same value of those endogenous causes. This condition is easily satisfied as long as I is properly exogenous. For instance, in the above example, we are assuming that how much water you give the plant is a random intervention, uncorrelated with factors like the weather. But if, as in fig. 5, I is somehow connected with the causes of X—for instance, if you only water then plants when it is sunny—things are trickier. The danger then is that those causes may independently affect Y, and any correlation observed between changes in I and changes in Y (in this example, changes in watering regime and changes in growth) cannot be assumed to reflect the causal influence of X on Y.[11]

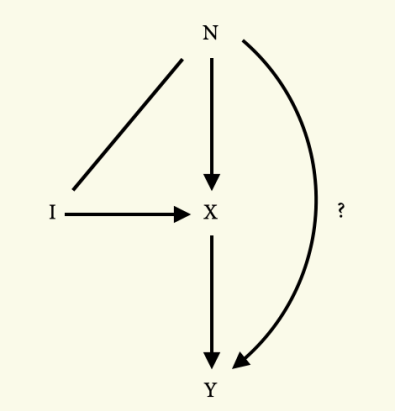

Figure 5: 'Confounded' soft intervention on X

Hence, if I is not a fully exogenous variable, what we need to consider is not some putative actual event of $I = i$, from which we might 'backtrack' to its normal causes N, but rather a *counter*factual event in which I takes the value i independently of its usual causes and correlates. Thus, the relevant information about hypothetical Interventions is strongly counterfactual, in the sense that it cannot be inferred from observable patterns of dependencies in the actual world, *including*

[11]Note that this last situation would involve a violation of condition I4. Thus someone might object that it is not actually a problem, given that this is already ruled out by the definition of a (soft) Intervention. However, remember that what we are concerned with here is not just an extensionally adequate analysis of causal claims, but one which might be somehow manifested in causal understanding, and on which ordinary actions can lead to causal knowledge. And the problem is that if I is not a properly exogenous event, then interventions involving I cannot, *in virtue of their causal history alone*, rule out any such I4-violating 'backdoor' influence on Y.

dependencies involving values of I. That is, if there is an association between I and the causes of X, the question, 'What would happen if we did I=*i*, given the actual causes of X', is not one that can be answered simply by noting what happens when we do I=*i*, because in that case the causes of X will be different from their actual values in the case where we do not do I=*i*. Such a situation would therefore undermine one of the principal advantages of interventionism, its ability to demystify causal counterfactuals by relating them to actual patterns of action and outcome that we can observe in the real world. In the next section I argue that this is in fact the case regarding our interactions with one another.

4 Entangled histories

Earlier I made the observation that our actions towards one another are typically not arbitrary, but rather arise, semi-endogenously, out of a background of mutual understanding and recognition, from general beliefs about the world to subtle and finely-tuned perceptual-motor skills. Why is this a problem? If the background conditions in question, such as the background beliefs of both participants to a conversation, count both as variable-level causes of one person's communicative actions, and as independent causes of their partner's responses, then we have a potential violation of condition I4, as in fig. 5.

Now, not all background causes of a phenomenon should be classified as variable-level causes. The presence of oxygen in the atmosphere is in some sense a background cause of my walking to the shops, but whether there is oxygen in the atmosphere should not be considered a variable-level cause of whether I walk to the shops.[12] In general, background conditions will count as variable-level causes of communicative actions only if there is some systematic relationship between which background conditions obtain and the specific content of what you say and do. And this much is indeed the case here: for example, what you say to someone in the course of a conversation is determined partly by your assessment of what is an appropriate, or reasonable, or acceptable thing to say or do, in turn partly de-

[12]For discussions of the issue of variable choice, see (Campbell 2010; Woodward 2016).

termined by your background beliefs. At the same time, what your partner does in response to receiving the message depends systematically on what they make of it, which in turn is determined partly by their background beliefs, which are correlated with yours.

A proper Intervention (including a soft one) would thus have to be an action that overrode or bypassed this communicative background in such a way that its content was entirely unpredictable on the basis of what had gone before. (Recall the paradigm of an intervention is a randomised experiment.) Clearly, this is a fanciful requirement: we do not go around saying arbitrary things to one another willy-nilly just to see what happens. We typically say things which seem appropriate to the situation by our lights, and our words' having their intended psychological effect is typically dependent on their seeming appropriate to the situation by the other person's lights. And the problem is now that, since the same background features are causally relevant to what one person says and how the other responds, it cannot be ruled out that they cause the response independently of any difference made by the putative intervention.

Go back again to Directions to the station. The point about it being a soft intervention was that its effectiveness depends on the other person being in various ways appropriately receptive, such as understanding what you say and being willing to accept it at face value. On Campbell's proposed amendment, the way around this problem is that we are not interested just in the association between the new belief and the outcome action (since this association could, for instance, be the upshot of an upstream common cause), but in the difference between what the person actually did, and what they would have done had you given them some other belief, holding the other features of the conversational background fixed.

The problem we now confront is that your action was no less a product of the conversational context than the person's response. That you pointed in one direction rather than another is not at all arbitrary: you are responding directly to the person's request for help, in doing so drawing on your shared understanding of the situation and your knowledge of the world. If you were to explain your action, you would say "They wanted to know where the station is, and it's over there, so

I pointed that way." From this perspective, your action looks entirely predictable. Thus, in imagining a situation in which the conversational background is the same and yet you act differently, we have to suppose the causal pathways to be somehow different from those at work in the actual world. The question is how we go about doing this while preserving those features of the situation that are relevant for the causal claims in question, and how we decide exactly which features to hold fixed and which to change. For instance, one way it might come about that you point east rather than west is if you have a different belief about where the station is. But then we have to ask where that different belief came from, and to suppose an entire alternative causal history to the interaction. And unless we simply assume from the outset that the person's new belief is a cause of their running, nothing in the causal setup of the situation rules out that, in this alternative history, something in the other person leads them to run off in what happens to be the direction you point, independently of any belief you give them.

It is no use objecting against the 'backtracking' counterfactual reasoning here. The point about insisting on an exogenous intervention is that it obviates the need for an explicit ban on backtracking (and the attendant questions how to secure this without recourse to Lewisian gerrymandering.) But when interventions are not suitably exogenous, as in this example, we face anew the same old problem for all difference-making theories of causation, namely how to specify the relevant class of hypothetical contrast cases.

4.1 Interventions as an ideal

However, there is a different way of looking at the situation. Woodward makes the suggestion that, when Interventions are not carried out, the notion of an Intervention can still function as a 'regulative ideal' in causal reasoning. (Woodward 2003, pp. 130–133) Here is how this could work. Even though, as a matter of fact, you were motivated to be helpful and to point to where you believed the station to be, there is a clear enough sense in which you could easily have decided to be unhelpful and point in the opposite direction. Focusing on this possibility, your action can be viewed as the product of a power of arbitrary choice—in F. P. Ramsay's (Ramsey

1926) phrase, an 'ultimate contingency'—even if the pattern of your actions in the actual world is far from arbitrary. We can think of this arbitrary power—which, following (Meek and Glymour 1994), we might call the Will—as functioning like a randomising device, rendering your choice statistically independent of its normal causes. The difference between these situations is depicted in figs. 6 and 7. From this perspective, it might be suggested, one can 'consider one's action as an Intervention', even when, as a matter of its actual causal history, it was no such thing.[13]

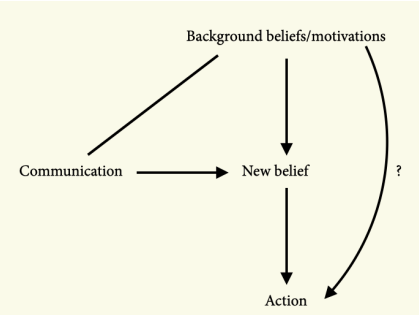

Figure 6: Ordinary communication

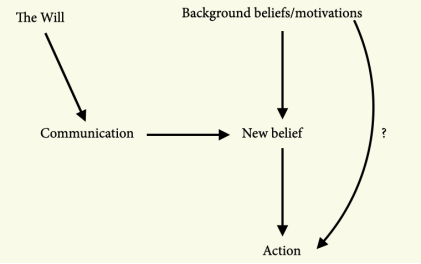

Figure 7: Intervention by the Will

In this mode of reasoning, it does not matter that your action was not, as a matter of fact, a random Intervention. What matters is rather the way in which the notion of an Intervention figures in guiding the construction of a counterfactual scenario. The idea is that, in understanding the other person's belief as a cause of their action, you are implicitly envisioning a situation in which your decision is the result of a kind of randomising mechanism—a different causal structure from the one that actually obtains. This is how it is possible to imagine you to have acted differently while holding all the other relevant features fixed. Note that, on this way of implementing Woodward's suggestion, there is no need for the hypothetical Intervention to be surgical: you suppose yourself to give them a new belief in the ordinary way, by giving them reasons for it. The point is that, by imagining your

[13]It is important to stress just how deviant and unusual the kind of event we are considering would have to be in order to play the relevant suppositional role. In particular, lying and other forms of garden-variety falsehood-telling to do not give us a good prototype for it. Lying is usually motivated. Thus, in a normal situation when you knowingly give someone misleading directions to the train station, you would have some reason for doing so, and so in reasoning from that supposition we are once again embroiled in backtracking.

action as produced by a power of arbitrary choice, it seems reasonably straightforward to imagine the relevant belief altered while keeping all other relevant features of their psychology fixed at their actual values.

However, there is a question of epistemic and conceptual priority here. Assuming that the above scenario is indeed possible, it is plausible to suppose that, had you performed that Intervention, there would have been a correspondingly different outcome. Thus, the causal facts about the case, combined with the possibility of soft Intervention, do imply and warrant the corresponding interventionist counterfactuals. But this does not establish that being able to draw this specific consequence of the causal facts has any special or canonical role in one's appreciation of those facts. On the contrary, it seems plausible that fantasising about the consequences of random unmotivated falsehood-telling is parasitic on the more basic cognitive achievement of grasping the causal role of belief in communicative interaction.

A more fundamental question is why our way of reasoning about this situation should be guided by the counterfactual supposition of arbitrary choice. To put it bluntly: if our actions are not ultimate contingencies, why should we consider them as though they were? Since our actions towards one another are in fact not at all random or arbitrary, it is obscure why the supposition that they are should play such a distinguished role in our hypothetical reasoning. In our interactions with the natural, non-human world to which we hold a basically detached or disinterested attitude, it is perhaps plausible to take this as in some sense the default, regarding any confounding correlations between our actions and the things we are trying to influence as accidental imperfections, to be eliminated or compensated for as a matter of scientific good practice. But when it comes to our ordinary understanding of one another, it is far from clear why this attitude should be desirable or appropriate.

It is worth stressing that, in expressing scepticism that the notion of an Intervention functions as a regulative ideal in our reasoning about our interactions with one another, we need not deny that there is an intimate connection between the appreciation of causal claims and grasp of the corresponding counterfactuals. To be sure,

in the above example, you would readily understand the causal structure of the situation to support counterfactuals such as: had I pointed the other way, they would have believed the station was to the east, and so would have run east rather than west. These counterfactuals have a non-backtracking interpretation: in evaluating them, we do not concern ourselves with questions of why, in the counterfactual situation, you pointed the other way. But this is not to say that the way we assess these counterfactuals is through envisaging, even implicitly, some well-defined hypothetical event in which you bring about the relevant change while leaving all the other features intact.[14]

5 Taking stock

It is worth reminding the reader at this point what the official problem is. The reason for insisting that interventions be independent of other background features is that we cannot rule out those features exerting an influence on the outcome variable, independently of the target variable. Now it may seem that this is not a realistic problem. In the case of the directions to the station, for instance, we would not ordinarily take seriously the possibility that the person's desire to reach the station plus their general background knowledge could cause them to run off to the west, without also the belief that the station is that way. We think there has to be an instrumental belief with a specific content in the explanation of their action. So one might think that the problem here is spurious.

The problem is indeed spurious, in the sense that it is not one we actually confront in our commerce with one another. It is, nevertheless, a problem that arises when we try to capture causal understanding purely in terms of correlations under (actual or hypothetical) interventions. The right conclusion to draw is that in a case like Directions to the station, you do not need your action to be an In-

[14]The fundamental point here is that (non-backtracking) counterfactual reasoning involves a special operation of *supposition*, in which—roughly speaking—the prior history of the events mentioned in the antecedent gets 'held fixed' when evaluating the consequent. Interventionism aims to capture this in terms of hypothetical reasoning about a special kind of *event*, whose causal history obviates the need for any distinctive mode of supposition. Scepticism that the interventionist strategy works in a given case does not amount to scepticism about the counterfactuals themselves. See (Joyce 2010) for a formal elaboration of this point.

tervention, or even to consider hypothetical ideal Interventions, in order for it to manifest causal psychological understanding. You are able to recognise your action as having affected that person's behaviour via a change in their belief, not because you have isolated that belief from its background causal context, but because you have a prior grip on the kinds of causal pathways by which we affect one another. For instance, in this case, in order to understand your action as a communicative intervention (not an Intervention) at all, you must appreciate that it works by giving the other person what is, in the context, a reason to think the station is to the west, which, given their aims, is a reason to run to the west. This is an instance of the more general idea, which we rely on constantly, that actions are caused by states that somehow rationalise or make sense of them, and hence that we can influence one another's behaviour by giving reasons for or against a course of action.

It is worth emphasising that the argument here does not turn on any kind of strong constitutive holism about the mind. I am not assuming that there is any *logical* inconsistency in the hypothesis that a bit of behaviour was caused independently of any rationalising belief. The point is just that, in our ordinary dealings with one another, we perforce make certain causal assumptions about the route by which our actions affect other people's conduct. But these are assumption which we do not, and perhaps could not, have independent evidence for in the form of information about actual or hypothetical Interventions on that causal route.

What is crucial here that the assumptions in question concern not only how the intervention variable (e.g. a communicative action) affects the target variable (e.g. a belief), as specified in the condition I1–4, but the whole route by which the target variable affects the outcome variable (e.g. an intentional action). For instance, in Directions to the station, your only reason for thinking that the influence of your verbal interventions on someone's behaviour is not confounded by a lurking extra variable is just that when you see them run off towards the station you are convinced that, in that context, their reaction has to be caused by a belief that corresponds to the content of your utterance and rationalises that behaviour. For this reason, it is not like the kinds of assumptions, familiar from causal modelling approaches, that could in principle be independently checked by further observations

and interventions. Rather, a grasp of the whole causal route of giving reasons for and against is more like a precondition of affecting one another intentionally via ordinary communication at all. Of course, if we could somehow surgically tweak someone's beliefs one at a time at random and compare the results with those of ordinary communicative interventions, we would then be in a position straightforwardly to verify, in interventionist terms, our causal assumptions about how communication works. But this is exactly the kind of controlled information we lack in our ordinary understanding.

The argument so far might be taken to incline towards scepticism that folk psychology is properly causal at all; according to this thought, genuinely causal understanding of ourselves belongs to the scientifically disinterested setting of lab-controlled or randomised experiments, in contrast to the engaged, reciprocal character of ordinary interpersonal understanding. In the remainder, I want to sketch a different way of thinking about causation in folk psychology and how it might be vindicated.

6 Intelligible connections and thick causal concepts

What interventionism plausibly gets right is that it is in the course of communicative interaction, not mere passive observation, that we come to understand one another as causally complex entitites, and to identify one another's causal levers in terms of the folk-psychological concepts of belief, desire, and all the rest. The problem is that our mode of learning through interacting does not fit well into the mould of inferring causal structure by eliminating, or assuming absent, possible confounders. On the contrary, it is only against the 'confounding' background of our shared world, and our natural sympathy with one another, that folk-psychological understanding is possible at all. We need to see how this natural sympathy might be, rather than an obstacle to genuine causal knowledge, perhaps to be overcome by science, a form of special insight into the workings of the mind.

The alternative picture I wish to sketch is one on which many of our folk-psychological concepts are instances of 'special' or 'thick' causal concepts: a

concept of a specific mode of causal influence which does not factorise into a general-purpose notion plus a restriction to a specific domain. Here is Nancy Cartwright's explanation of the idea:

> All thick causal concepts imply 'cause'. They also imply a number of noncausal facts. But this does *not* means that 'cause' + the noncausal claims + (perhaps) something else implies the thick concept. For instance we can admit that *compressing* implies *causing* $+x$, but that does not ensure that *causing* $+x + y$ implies compressing for some non-circular y. (Cartwright 2004)[15]

In other words, thick causal concepts involve not just a cause of a certain type bringing about an effect of a certain type, but its doing so in a specific *way*—and, although we may be able to give a more or less detailed characterisation of that way and its typical features, we may not be able to give necessary and sufficient conditions for it except trivially, by employing that very causal concept.[16]

To see how this suggestions works in the psychological case, let us fix on an example that offers a somewhat different paradigm of what we do to each other than the example of the directions to the station: namely, interactions involving joint attention to a shared environment.

CHURCH FAÇADE Strolling around town with a companion, you point out an interesting figure on a church façade, and they make an appreciative murmur. You have brought it about that their visual attention is directed, with yours, towards the figure, which, given their general likes and sensibilities, causes a certain reaction in which they outwardly express their aesthetic pleasure.

Your action here is not an actual Intervention because it arises out of the same context of mutual awareness and shared interest, and is caused by the very same object and its properties of the figure, that cause their reaction. Fig. 8 is an attempt to depict the main lines of this causally complex situation. The problem (or pseudo-problem) discussed in connection in the previous case is then that your

[15]The idea of thick causal concepts is generally attributed to G. E. M. Anscombe's seminal essay 'Causality and Determination' (Anscombe 1981). Another important source of inspiration is Bernard Williams's (Williams 1985) discussion of thick ethical concepts.

[16]For discussions of non-factorisability in other domains, such as knowledge and intentional action, see (Ford 2008; Williamson 2000).

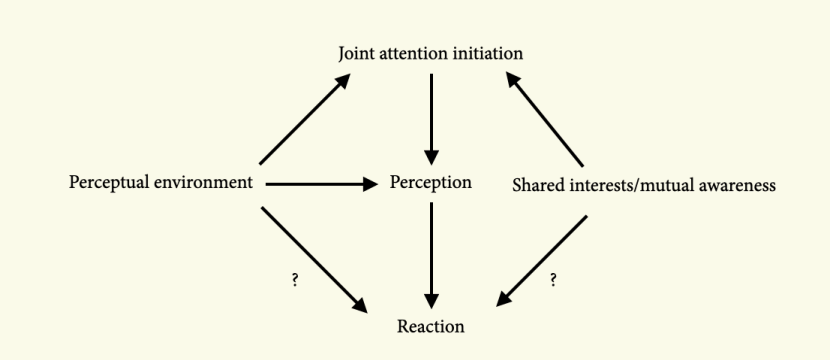

Figure 8: Joint attention

action does not appear to rule out that your companion's reaction is caused by these background features of the context, independently of your directing their perceptual attention to the object. And this seems if anything even harder to take seriously than the possibility that someone's behaviour might be caused independently of any new belief. It is just not possible for the mere presence of the object, the absence of perceptual obstacles, and their background dispositions, to have caused that response: they have to be actively and consciously attending to it in order to respond in that way. It is almost as if you can 'see' the causality in the other person turning their gaze and responding to what they see—just as it seems you can 'see' the transfer of motion when one billiard ball strikes another.

I do not want to put much weight on the idea that psychological causality is literally observable. Nevertheless, we should take seriously the thought that, in joint attentional interactions, the triangular causal structure of the situation can be epistemically open to both participants, and so open partly in virtue of their shared sensitivity to the object of attention and sympathy with one another. A participant in a joint attention interaction is able to recognise both theirs and the other person's actions *as* responses causally mediated by perceptual attention to the object, and are able to exploit this understanding to elicit different responses from the other person by directing their attention to different aspects of the perceptual scene.

There are various aspects to this. One is of course an appreciation of some of the basic mechanics of the perceptual processes involved, such as the need for

a clear line of sight for vision.[17] Capturing the role played by attention, beyond simple perceptibility, however, involves more than this. Participants to joint attention understand one another's behaviour to be causally sensitive to the attended object in a very special way—as a fitting or appropriate response to the object's features, which the other produces precisely because they are consciously attending to those very features. This can be seen in the fact that, if someone responds in a surprising way to to a jointly attended object or scene, a typical response is not merely to take the contingency as evidence for a causal connection, but rather to try to understand what might have made that response appropriate. For instance, if someone points and laughs when there is nothing obviously amusing to see, you might yourself try to discern what is funny in the situation so that you can share their response, rather than simply adding this idiosyncratic contingency to your general knowledge of the world. In this respect, the causal understanding inherent to joint attention involves, in addition to the basic mechanics of perceptual causality, the more normatively laden idea of environmental features as calling out for certain evaluative or affective responses in subjects whose perceptual attention is directed at them.[18]

This gloss on the paradigm of joint attention sees the normative and psychological as intertwined in a way that stands in contrast with a more conventional philosophical picture. On a standard construal, causally explaining a psychological occurrence is basically a separate matter from assessing its normative standing: if normative considerations come in at all at the level of causal explanation, it is by way of providing global side-constraints on what combinations of psychological attributions are intelligible. By contrast, in the above case, the normative dimension comes in by way of characterising a specific, local channel of causal influence: that of a fitting or appropriate response to a perceptible feature. That is, the causal understanding involved in joint attention involves the idea that there is a specific way in which your companion's perceptual attention causes their response to the object, one that involves their perceptually recognising the object as meriting a

[17] This already raises interesting questions about the extent to which a grasp of the spatial constraints on perception here can be adequately captured in interventionist terms; cf. (Roessler 2011).

[18] For a rich discussion of the normative dimensions of joint attention, see (Eilan et al. 2005).

certain response in virtue of how it looks.

I propose that the distinctive fusion of the causal and the normative we see in the understanding of joint attention exhibits exactly the structure of explanations in terms of thick causal concepts. The normativity inherent to the idea of appropriately responding to an object's perceptibly manifest features is not a separate component that can be factorised out of the causal relation between an object's features and a subject's state of perceptual attention, or between perceptual attention and a behavioural response, but rather characterises the specific manner or mode in which an object brings about a behavioural response in a suitably attuned and perceptually attentive subject.

This explanatory schema stands in marked contrast to an interventionist account of causal understanding in two, related ways. First, the appeal to thick causal concepts allows us to resist the idea that the canonical evidence for all causal explanations has to take the form of information about actual or hypothetical Interventions. Although claims framed in terms of thick causal concepts may have have implications for what would happen in various counterfactual circumstances, including those in which events occur that are Interventions, those implications need not play any specially distinguished role in the understanding of the claims.

Secondly, and more radically, the ecological background of a shared environment and sensibility comes in here, not as an obstacle or confounder to be overcome or compensated for, but as part of the basis of the grasp of the relevant causal concepts. In understanding someone's reaction of revulsion to something disgusting or offensive, for instance, the understanding of the specific way in which their reaction came about involves my being able to grasp why that reaction was appropriate, which ultimately depends on my sharing enough of their sensibilities for the reaction to be comprehensible to me.

So far I have characterised just one component of our folk-psychological causal understanding, the way in which we understand another person's responses to a perceptible object in joint attention. The difficult question for the approach just sketched is how it generalises to other aspects of the understanding of psychological causality; in particular, how it might apply to understanding the causal role of belief,

for instance as manifest in the earlier example of the directions to the station. There is an intuitive sense in which the causal role of beliefs, especially beliefs about a causally remote subject-matter, is less 'observable' than that of states like perceptual attention. Moreover, the causal role of belief is notoriously multifarious and open-ended, with no way of specifying finitely the ways in which a belief of a given type might come about, or the new beliefs or actions it might give rise to. So one might think there are good grounds to doubt that the idea of a certain specific type of interaction, picked out by a given thick causal concept, really applies here.

What the above complaint gets right is that a grasp of the causal role of belief is clearly a much more sophisticated achievement than the ability to elicit and understand responses from another person by directing their perceptual attention. Nevertheless, I think we can recognise something like the same schematic explanatory structure at work in the understanding of the causality of belief in cases like Directions to the station. The point there, which was the chief sticking point for an interventionist approach, is that in that situation you have to be at least as confident about the specific way in which you have brought about a behavioural change as you are in any of the other causal facts: you have caused the person to run off to the west by giving them a belief that the train station is that way, and thereby, given that they want to get to the train station, giving them a reason to run over there. In other words, you must recognise that this is an instance of affecting someone's conduct by giving them reasons for or against a course of action; and your being able to recognise this is partly a matter of your sharing the same world, and the same sense of what is a reason for what. The ability to give and understand reasons for action, and to recognise when someone has acted on a certain reason, is, on this proposal, part of what constitutes our grasp of the 'causal role of belief': not a list of possible causal inputs and outputs, but the ability to initiate, and to recognise, a certain normatively laden pathway of causal influence.

Clearly, there is much more to be said on this topic than I can attempt here. The ability to give and recognise belief-dependent reasons for action is not a single or simple achievement, and it is an interesting and fruitful question just how it is related, both conceptually and developmentally, to simpler abilities for tracking

causal threads in the mind that are, so to speak, nearer the surface. The possibility I am raising here is just that, when it comes to the causal component of belief-involving psychological explanations, we are not applying a general-purpose notion of causal influence to a certain domain, but rather are deploying specific, thick concepts which are tailor-made to our understanding of one another as psychological beings.

7 Conclusion

In this paper I have been principally discussing our causal psychological concepts. One might have the worry that this does not tell us very much about the phenomenon of psychological causation itself. Isn't it possible that our concepts will turn out to be ill-grounded, and fail to track any causal relations that genuinely obtain in the mind?

From this perspective, the value of the special causal concepts proposal is that it offers an alternative, deflationary way of conceiving how the epistemology of psychological causation relates to its metaphysics. A standard philosophical approach to mental causation takes it as read that folk psychology makes causal claims, but takes spelling out the commitments of those claims to be the task of a general philosophical theory of causation. The tendency is therefore to open up a rift between ordinary understanding and the official causal theory, making it hard to see how folk psychology could amount to anything more than an agglomeration of dubious causal hunches. Interventionism promises to do better on this score by relating causal claims systematically to ordinary patterns of action and inference; but, I have argued, it founders in the details when we look more closely at the causal structure of ordinary communicative actions.

The prospect raised by the special causal concepts proposal is, by contrast, that there may be no deeper or more accurate way to describe causation in the mind than simply in the very terms of folk psychology, which are already saturated with causality. We are equipped with a rich array of interrelated concepts in terms of which to understand the pushes and pulls of the inner life; these are the best means

we have for tracking causal relations between psychological occurrences, and they resist translation into a less idiosyncratic, general-purpose theory of causation. In the philosophy of mind it remains a live option to take this ordinary understanding at face value, as a special and untranslatable form of causal knowledge of the psychological domain.

References

Anscombe, G. E. M. (1981). 'Causality and Determination'. In: *Metaphysics and the Philosophy of Mind*. orig. 1971. University of Minnesota Press.

Apperly, Ian (2010). *Mindreaders: The Cognitive Basis of "Theory of Mind"*. Psychology Press.

Campbell, John (2006). 'An Interventionist Approach to Causation in Psychology'. In: *Causal Learning: Psychology, Philosophy and Computation*. Ed. by Alison Gopnik and Larry J. Schulz. Oxford University Press, pp. 58–66.

— (2008). 'Causation in Psychiatry'. In: *Philosophical Issues in Psychiatry: Explanation, Phenomenology, and Nosology*. Ed. by Kenneth S. Kendler and Josef Parnas. Johns Hopkins University Press.

— (2010). 'Control Variables and Mental Causation'. In: *Proceedings of the Aristotelian Society* 110.1pt1, pp. 15–30.

— (2020). *Causation in Psychology*. Harvard University Press.

Carruthers, Peter and Peter K. Smith (1996). *Theories of Theories of Mind*. Cambridge University Press.

Cartwright, Nancy (2004). 'Causation: One Word, Many Things'. In: *Philosophy of Science* 71.5, pp. 805–819.

Dennett, Daniel C. (1981). 'True Believers : The Intentional Strategy and Why It Works'. In: *Scientific Explanation: Papers Based on Herbert Spencer Lectures Given in the University of Oxford*. Ed. by A. F. Heath. Clarendon Press, pp. 150–167.

— (1991). 'Real Patterns'. In: *Journal of Philosophy* 88.1, pp. 27–51.

Dennett, Daniel Clement (1981). *The Intentional Stance*. MIT Press.

Eilan, Naomi et al. (2005). 'Joint Attention and the Problem of Other Minds'. In: *Joint Attention: Communication and Other Minds: Issues in Philosophy and Psychology*. Oxford: Clarendon Press.

Ford, Anton (2008). 'Action and Generality'. PhD thesis. University of Pittsburgh.

Geertz, Clifford (1973). *The Interpretation of Cultures*. Basic Books.

Goldman, Alvin L. (2008). *Simulating Minds: The Philosophy, Psychology, and Neuroscience of Mindreading*. Oup Usa.

Gopnik, Alison, Clark Glymour et al. (2004). 'A Theory of Causal Learning in Children: Causal Maps and Bayes Nets'. In: *Psychological Review* 111.1, pp. 3–32.

Gopnik, Alison and Tamar Kushnir (2003). 'Inferring hidden causes'. In: *Proceedings of the 25th Annual Meeting of the Cognitive Science Society*. Ed. by R. Alterman and D. Kirsh. Boston.

Gopnik, Alison, Tamar Kushnir and Laura Schulz (2007). 'Learning from Doing: Intervention and Causal Influence'. In: *Causal Learning: Psychology, Philosophy, and Computation*. Oxford University Press.

Gopnik, Alison and Andrew N. Meltzoff (1997). *Words, Thoughts, and Theories*. MIT Press.

Gopnik, Alison and Laura Schulz (2007). *Causal Learning: Psychology, Philosophy, and Computation*. Oxford University Press.

Hall, Ned (2000). 'Causation and the Price of Transitivity'. In: *Journal of Philosophy* 97.4, p. 198. DOI: 10.2307/2678390.

Halpern, Joseph Y. and Judea Pearl (2005). 'Causes and Explanations: A Structural-Model Approach. Part I: Causes'. In: *British Journal for the Philosophy of Science* 56.4, pp. 843–887.

Hitchcock, Christopher (2001). 'The Intransitivity of Causation Revealed in Equations and Graphs'. In: *Journal of Philosophy* 98.6, pp. 273–299.

Hoerl, Christoph (2011). 'Perception, Causal Understanding, and Locality'. In: *Perception, Causation, and Objectivity*. Ed. by Johannes Roessler, Hemdat Lerman and Naomi Eilan. Oxford: Oxford University Press, pp. 207–228.

Hoerl, Christoph, Teresa McCormack and Sarah R. Beck (2011). *Understanding Counterfactuals, Understanding Causation: Issues in Philosophy and Psychology*. Oxford:: Oxford University Press.

Joyce, James M. (2010). 'Causal Reasoning and Backtracking'. In: *Philosophical Studies* 147.1, pp. 139–154.

Kaiserman, Alex (2020). 'Interventionism and Mental Surgery'. In: *Erkenntnis* 85.4, pp. 919–935.

Kenny, A. (1963). *Action, Emotion And Will*. Ny: Humanities Press.

Lewis, David (1987). 'Causation (with Postscripts)'. In: *Philosophical Papers, Vol. 2*. orig. 1973. New York, Oxford: Oxford University Press.

McCormack, Teresa, Christoph Hoerl and Stephen Butterfill (2011). *Tool Use and Causal Cognition*. Oxford University Press.

Meek, Christopher and Clark Glymour (1994). 'Conditioning and Intervening'. In: *British Journal for the Philosophy of Science* 45.4, pp. 1001–1021.

Melden, Abraham I. (1961). *Free Action*. Routledge.

Nichols, Shaun and Stephen P. Stich (2003). *Mindreading: An Integrated Account of Pretence, Self-Awareness, and Understanding Other Minds*. Oxford University Press.

Pearl, Judea (2009). *Causality: Models, Reasoning and Inference*. 2nd ed. orig. 2000. Cambridge University Press.

Ramsey, F. P. (1926). 'Truth and Probability'. In: *Philosophy of Probability: Contemporary Readings*. Ed. by Antony Eagle. Routledge, pp. 52–94.

Roessler, Johannes (2011). 'Perceptual Causality, Counterfactuals, and Special Causal Concepts'. In: *Understanding Counterfactuals, Understanding Causation*. Ed. by Christoph Hoerl, Teresa McCormack and Sarah R. Beck. Oxford University Press.

Sperber, Dan, David Premack and Ann James Premack (1995). *Causal Cognition: A Multidisciplinary Debate*. Oxford University Press UK.

Spirtes, Peter et al. (2000). *Causation, Prediction, and Search*. 2nd ed. orig. 1993. Mit Press: Cambridge.

Taylor, Charles (1971). 'Interpretation and the Sciences of Man'. In: *Review of Metaphysics* 25.1, pp. 3–51.

Waldmann, Michael R., ed. (2017). *The Oxford Handbook of Causal Reasoning*. Oxford University Press.

Williams, Bernard (1985). *Ethics and the Limits of Philosophy*. Harvard University Press.

Williamson, Timothy (2000). *Knowledge and its Limits*. Oxford: Oxford University Press.

Winch, Peter (1958). *The Idea of a Social Science and its Relation to Philosophy*. Routledge.

Woodward, James (2003). *Making Things Happen: A Theory of Causal Explanation*. Oxford University Press.

— (2007). 'Interventionist Theories of Causation in Psychological Perspective'. In: *Causal Learning: Psychology, Philosophy, and Computation*. Oxford University Press.

— (2016). 'The Problem of Variable Choice'. In: *Synthese* 193.4, pp. 1047–1072.

— (2021). *Causation with a Human Face*. Oxford University Press.