Erkenn (2007) 67:321–336
DOI 10.1007/s10670-007-9074-x

ORIGINAL ARTICLE

Intentional Systems Theory, Mental Causation and Empathic Resonance

Marc V. P. Slors

Received: 30 July 2006 / Accepted: 30 April 2007 / Published online: 11 August 2007

Abstract In the first section of this paper I argue that the main reason why Daniel Dennett's Intentional Systems Theory (IST) has been perceived as behaviourist or antirealist is its inability to account for the causal efficacy of the mental. The rest of the paper is devoted to the claim that by emending the theory with a phenomenon called 'empathic resonance' (ER), it can account for the various explananda in the mental causation debate. Thus, IST + ER is a much more viable option than IST, even though IST + ER assigns a crucial role to the phenomenology of agency, a role that is incompatible with Dennett's writings on consciousness.

The most fundamental thesis of Daniel Dennett's intentional systems theory (IST; Dennett 1978, 1987) is that the ontology of mental states cannot be considered in abstraction from the epistemology of mental state ascription. From this thesis, a number of attractive features follow. IST respects the distinction between the sub-personal and the personal level of description (the distinction is Dennett's (1969) own). It resists the reification of beliefs and desires like no other theory that aspires to a form of realism about the mental. It does not imply theses about the nature of the brain that may or do contradict the findings of neuroscience, nor does it need to postulate theses about the brain that are immune to empirical investigation.

But IST has never been an overly popular position. This is due mainly to its perceived behaviourist, instrumentalist character—its failure to secure mental realism, despite its aspirations. One way to understand this opposition against IST, as I shall explain in the next section, is to construe it as a worry about the inability of IST to accommodate the phenomenon of mental causation. Indeed, Dennett does not even attempt to make room for mental causation within IST.

M. V. P. Slors (✉)
Department of Philosophy, Radboud University Nijmegen, P.O. Box 9103, Nijmegen 6500 HD, The Netherlands
e-mail: marc.slors@phil.ru.nl

This is unfortunate. For it appears to suggest that the attractive features of IST are only to be had when we entirely, and according to most philosophers unrealistically, give up on mental causation. In this paper I address the question whether it is possible to retain IST while at the same time making an emendation to it so that mental causation can be accommodated by the theory. I will defend a positive answer: if IST is combined with relatively recent insights into a phenomenon referred to as 'empathic resonance', I argue, it can make room for a form of mental causation that suffices to satisfy our intuitions.

The paper is set up as follows: In the next section, I outline IST and interpret the anti-realism charge in terms of the failure to accommodate mental causation. In Sect. 2, I shall introduce empathic resonance. In Sect. 3, I will combine empathic resonance with IST. This combination, I argue in Sect. 4, makes room for 'phenomenal causation', the causal efficacy of phenomenal states. (This result is incompatible with Dennett's theory of consciousness. However, I will emphasize that IST itself is independent of this theory of consciousness). In Sect. 5, I shall argue that in the proposed theory propositional attitudes cannot be considered causally efficacious, though they *are* what Jackson and Pettit have called 'causally relevant'. In Sect. 6, I argue that the combination of phenomenal causation and the causal relevance of propositional attitudes suffice to account for the various exlananda that are discussed in the debate on mental causation.

1 Intentional Systems Theory, Mental Realism and Mental Causation

IST takes its cues from Wittgenstein, Ryle and Quine. From Wittgenstein and Ryle it extracts the idea that systems really *are* intentional or 'minded' systems, if their behaviour can fruitfully be interpreted as issuing from beliefs and desires or in short: folk-psychological states. From Quine it takes the idea that folk-psychological interpretations of behaviour can be indeterminate—it is possible to have rival interpretations between which we cannot decide.

IST is a form of interpretationism; the ontology of beliefs and desires is not considered independent of their epistemology. As Dennett has put it (restricting himself to beliefs only in this quote):

> My thesis will be that while belief is a perfectly objective phenomenon (that apparently makes me a realist), it can be discerned only from the point of view of one who adopts a certain *predictive strategy* [the strategy of using folk-psychology; M.S.] and its existence can be confirmed only by an assessment of the success of that strategy (that apparently makes me an interpretationist) (Dennett 1987, p. 15).

When interpreting the behaviour of systems in terms of beliefs and desires, we are adopting what Dennett calls 'the intentional stance.' That is, we interpret behaviour in folk-psychological terms. Though we are inclined to adopt the intentional stance rather quickly, e.g. when interpreting fellow humans, we can always leave that stance and adopt a different one, such as the physical stance or the design stance.

It may now seem as if IST claims that *we* turn a system into an intentional one merely through interpretation. But that is incorrect. A system has beliefs and desires, according to IST, if these notions really *help* in understanding and predicting the behaviour of that system. 'Exhibiting behaviour that can fruitfully be interpreted using the intentional stance' is a property of systems that cannot be assigned by us at will. It is an objective (though relational) property.

This is why Dennett considers himself a mental realist—albeit a 'mild realist' (Dennett 1987, pp. 69–81, 1991a). This realism-cum-interpretationism is well captured by his comparison of intentional states in (folk-)psychology with centres of gravity in Newtonian mechanics. Like intentional states, a centre of gravity is a notion that abstracts away from the concrete 'behaviour' of an object, but it allows us to describe and predict the 'behaviour' of that object with a great degree of accuracy. The notion of intentional states, like the notion of a centre of gravity, is an informative notion that discloses information about the world that cannot be accessed in any other way.

But neither mental states, nor centres of gravity are real *entities* or *objects*. Dennett makes an effort to combine a strong resistance against the reification of beliefs and desires with what he considers a form of mental realism. Beliefs are real, he claims. But to consider them entities that exist in our heads would be to apply a concept at the sub-personal level of description that has its use (and hence in a Wittgensteinian view, it's meaning) at the personal level of description only. Denying that beliefs are entities in the head is not denying their reality, it is to make a claim about their nature: they are *states* of whole persons (or other systems), not *parts* of them. People *really* believe things. Just like people can *really* be fatigued. But 'a belief' is just as much a *thing* as 'a fatigue'.

To many philosophers, including myself, this strong resistance to reification of belief and this strict separation of the personal and sub personal level of description are attractive features of IST. But at the same time, this position is generally viewed as not succeeding in establishing a mental realism that is strong enough to cater for our commonsense needs. The perception is that IST is an 'as if' theory of mental state ascription (McCulloch 1990).

Dennett acknowledges this worry and has attempted to put it at ease at several occasions (most prominently in Dennett 1991a, but earlier in 1987, pp. 37–42 and pp. 69–81; part of the point of these texts gets a more elaborate treatment in Chapter 2 of his 2003a). Folk-psychological predicates track patterns in behaviour, he argues (just like centres of gravity track patterns in movement of objects). And these patterns may require a perceiver in order to be recognized, but that doesn't mean they are merely 'in the eye of the beholder'. They exist 'out there'. And if, as Dennett argues, our folk-psychology is the only access we have to these patterns, then folk-psychological predicates disclose a part of objective reality for us.

Although there is much more to say about this patterns-response, I shall set it aside. For although it does address some of the intuitive worries about the type of mental realism Dennett defends, it does not address all of them. Take Jerry Fodor's description of realism about propositional attitudes in general:

> I propose to say that someone is a *Realist* about propositional attitudes iff (a) he holds that there are mental states whose occurrences and interactions cause behaviour and do so, moreover, in ways that respect (at least to an approximation) the generalizations of common-sense belief/desire psychology; and (b) he holds that these same causally efficacious mental states are also semantically evaluable (Fodor 1985, p. 78).

From this quote, it is clear that Fodor takes causal efficacy to be a hallmark of the reality of intentional states. In doing so, he is in good company. Shoemaker (1980), for instance, argued influentially that the identity conditions of properties are to be spelled out in terms of their causal powers. It is impossible to discuss the connection between causal efficacy and reality in detail here. But I shall treat the connection as a very strong intuition that cannot be ignored.

Yet Dennett does seem to ignore it. The patterns response in no way addresses the issue of mental causation. Part of Fodor's charge (and that of many others) is that IST leaves no room for causally efficacious mental states. According to IST, beliefs are like centres of gravity. But centres of gravity lack causal efficacy too, and that is why many do not consider them real, *pace* Dennett. The failure to accommodate mental realism, then, can be construed as at least partly being due to a failure to accommodate mental causation.

In reply to this, the patterns response can in turn be construed as arguing that mental causation is *not* a precondition for mental realism. But that would raise the question why Dennett never addresses the issue of mental causation directly. And why he never bothered to reject Davidson's original argument for taking reasons to be causes (Davidson 1963). I submit that it is precisely because the patterns response ignores the issue of mental causation that most philosophers are not convinced by it.

To this situation, IST can react in two ways. One—which I take to be Dennett's option—is to argue against the idea that mental states must be construed as causes in order to be considered real. Another option would be to see whether some form of mental causation can be made compatible with IST's interpretationism. It is this latter option that I shall pursue.

2 Empathic Resonance

What I should like to propose is—to put it in a slogan-like way—that IST may leave room for a form of mental causation if the behavioural patterns that are interpreted using the intentional stance are not conceived of as consisting of bodily *movements*, but rather of bodily *gestures*. To put some flesh on these bare bones, what is required for this is that the behaviour that is being interpreted is not just passively perceived via ordinary sensory perception, but rather perceived via what I shall label 'empathic resonance'. In this section I shall first introduce empathic resonance. In the next section I shall distinguish it from and connect it to IST. In the rest of the paper I shall explain how the combined theory can deal with mental causation.

Empathic resonance is a phenomenon in the domain of social cognition that plays a significant role in a variety of theories of social cognition. It can be introduced as an extrapolation of the phenomenon of emotional contagion.

Emotional contagion is a "multiply determined family of social, psychophysiological, and behavioural phenomena" (Hatfield et al. 1994, p. 7) in which one person directly 'picks up' the emotional 'drive' behind the facial expression, bodily posture, gesture, etc. of another person. Consider the following example by Robert Gordon in which he describes Hermia's attempt to understand and predict the behaviour of Demetrius in a scene added to Shakespeare's Midsummer Nights Dream:

> (…) [T]he sight of Demetrius' facial expression would probably have produced a similar expression on Hermia's face—even if not a visually detectable expression, at least the corresponding pattern of muscular innervation. And these copy-cat innervation patterns, at least when they replicate another's expression of emotion, tend to produce an emotion in us, typically (where there are no relevant cultural differences), an emotion similar to the one that caused the other's original expression. Thus, by replicating the facial expressions of others, we would tend to 'catch' the emotions expressed (Gordon 1996, p. 13).

The example is fictional. The phenomenon described in it is very real (see e.g. Melzoff and Gopnik 1993) and very salient in everyday social interaction (see especially Hatfield et al. 1994, pp. 79–127).

In the most elaborate study into this phenomenon published so far, Hatfield et al. propose a plausible and hence widely accepted analysis of this phenomenon in two steps: (i) the first step is the mimicry of observed behaviour, usually involuntary, and often below the threshold of conscious awareness. The tendency to mimic behaviour, especially of the mother, is present in people directly from birth onwards (Melzoff and Moore 1977). The discovery of mirror neurons (Pellegrino et al. 1992; Gallese et al. 1996; Rizzolatti et al. 1996; Fadiga et al. 1995) is an important contribution (but probably not more than that) to our understanding of the neurological mechanisms behind such mimicry. (ii) The second step is derived from Darwin's (1872/1965) and James' (1890/1984) observation that emotional experience is profoundly affected by 'feedback' from facial muscles, the internal perception of visceral organs, the proprioception of bodily posture etc. (see also Damasio 1994). Emotions are enhanced by their corporeal expression to such an extent that the relevant muscular activity (or, as the discovery of mirror neurons now suggests, even the relevant premotor activity in the brain; Iacoboni 2003; Carr et al. 2003) when mimicked, would *produce* a faint echo of the emotion behind the behaviour that is being mimicked.

In order to turn this phenomenon into a primitive, pre-conceptual form of ascription of emotions, as Gordon and many others (e.g. Gallese and Goldman 1998) at least seem to suggest, a third step is needed. For experiencing an 'echo' of someone else's emotion is not yet attributing anything to anyone (see Goldman 2006, pp. 133–134). This third step may consist of employing a version of the theory

of mindreading (see e.g. Perner 1996). I shall defer issues of ascription to the next section.

Iacoboni (2003) speaks of 'empathic resonance' instead of 'emotional contagion'. That seems to reflect an appropriate broadening of the phenomenon to situations in which the meaning of the term 'emotion' would have to be stretched beyond recognition. One of the best examples of the phenomenon, for instance, is the inclination to yawn when one sees, hears or even reads about someone else yawning (Provine 1986, 1989). But unlike e.g. laughter (Provine 1992) the 'feel', 'urge', or 'drive' behind it falls outside most people's class of emotions. So, a more general term such as 'empathic resonance' seems appropriate. It signifies the pick up and often the implicit attribution of the 'feel/urge/drive' behind basic bodily 'actions' such as facial expressions, gestures, bodily postures expressive of intended behaviour etc.

In the case of yawning, the behaviour with which we tend to resonate is likely to be unintentional. But we also (and very often) resonate with intentional behaviour. And that is where empathic resonance becomes a form of social cognition. Think of the pick up of anger in someone's behaviour. But also of the way we directly 'perceive' someone's intention to open a door, pick up a glass, or shake hands. What we pick up in the case of intentional action is an 'urge' or 'drive' behind or in a sense *in* an action as an unconceptualized basic intention (but *not* (yet) as beliefs and desires; for elaborate overviews of the kind of social cognition that empathic resonance allows for, see e.g. Gallagher 2004, 2005, Ch. 9; and Bloom 2004, Ch. 1).

By replacing the term 'emotional contagion' with 'empathic resonance' and by recognizing that by means of such resonance we can pick-up on basic intentions of others, we may label the idea that basic intentions are 'visible' in gestures, bodily posture, and facial expressions, a 'Darwin–James-like view of intention-in-action' (to distinguish it from Searle (2001), whose notion of intention-in-action is about propositional attitudes rather than—often emotionally laden—basic intentions).

3 Empathic Resonance and Intentional System Theory Combined

When we adopt the intentional stance in order to interpret the behaviour of others and when we empathically resonate with the behaviour of someone else, superficially speaking we do similar things. In both cases a mental drive is postulated on the basis of observed behaviour. The similarity is, indeed, merely superficial, as I shall explain below. As a consequence, something needs to be said about a *division of labour* between the two when it comes to mental state ascription on the basis of observed behaviour. I will argue that the division of labour must be such that adopting the intentional stance in stereotypical cases involves interpreting information about the behaviour of others acquired via empathic resonance (ER).

The difference between ER and IST can best be elaborated on by assigning them a location on the map of possible options in the debate over the nature of mental state ascription. The debate is dominated by various forms of 'theory theory' (TT) and various forms of 'simulation theory' (ST), though Shaun Gallagher's (2004, 2005) interaction theory (IT) might be regarded as a third option. ER is entirely at home in

the camp of ST or IT; IST is a form of TT. Let me briefly say something about these claims.

It is certainly not the case that ER fits all forms of simulationism. In particular all forms of simulationism that involve a conscious effort to place oneself imaginatively in the shoes of the person whose behaviour is being interpreted (e.g. Harris 1991, 1992; Heal 1986, 1996) involve much more than just ER and do not even mention it. Other versions, such as Gordon's (1996) radical simulationism and to a lesser extent Perner's (1996) TT/ST mix-theory, explicitly involve something like ER as *part* of what is involved in mental state attribution. And then there is Goldman and Gallese's (1998) version of simulationism as backed up by mirror neurons that almost describes ST as a form of ER. Finally, ER is a *conditio sine qua non* for the majority of social cognition capacities described in Gallagher's IT.

More important than the fact that ER does appear at home in IT and many versions of ST is the fact that ER involves no theory use. It doesn't involve folk-psychology. Regardless of whether or not it is part of ST or IT, it is *not* a version of TT. IST, by contrast, explicitly *is* a version of TT. Attribution of intentional states is a form of theorizing according to Dennett, as e.g. the comparison between intentional states and centres of gravity shows. Moreover, Dennett is of the opinion that simulation ultimately collapses into theory use (see especially Dennett 1987, pp. 100–101).

So, ER and IST describe different 'activities', when it comes to the ascription of mental drives behind behaviour. And that poses the question of the division of labour between them. It might seem plausible to argue that ER and adopting the intentional stance serve related but different purposes and hence occur in different situations. But this would imply that we do not usually empathically resonate with behaviour we interpret using the intentional stance. And that seems typically wrong. We *do* often adopt the intentional stance after we empathically resonate with someone's behaviour. In fact, we typically start to interpret the behaviour of other people *after* we perceive intentionality in that behaviour; we do not interpret sneezes and hiccups using the intentional stance. The intuitive division of labour between ER and IST in stereotypical cases that I would propose is as follows: ER is used to determine whether the behaviour of the system warrants further intentional interpretation (not all behaviour we resonate with does: yawning doesn't require further interpretation (usually), aggressive gestures do); if so, the intentional stance is adopted in order to interpret the behaviour in folk-psychological terms. There is much to say about how this interpretation proceeds. But given that none of the following depends on these details I will leave it at this here.

This proposal implies a deviation from Dennett's original IST. For it implies that we typically adopt the intentional stance towards beings with which we resonate empathically, whereas Dennett (being a theory theorist) used to speak freely of us adopting the stance towards suspension bridges and thermostats. On my proposal, these would be atypical *as if* cases of applying the intentional stance that are derived from *real* cases.

Strikingly, Dennett makes a proposal similar to the above in his last book (Dennett 2006), thus indicating—but not explicitly recognizing—a change in his position. He does not speak of empathic resonance, but of our having and using an

'Agent Detection Device' (ADD) that is similar to what Simon Baron Cohen calls an 'Intentionality Detector' (Baron Cohen 1995). He describes it as "a Good Trick of evolution to discriminate banal motions (the rustling of leaves or swaying of seaweed) from those that signal the presence of a predator, prey, mate or rival conspecific" (2006, p. 108–189). To be sure, Dennett's ADD need not involve ER. The similarity between Dennett's move and my proposal is not so much in a possible overlap between ADD and ER (though I strongly believe there is overlap at least in the human case), but in the fact that there appears to be agreement over the idea that the job of judging a system suitable for intentional interpretation is *not* (or: no longer) assigned to the intentional stance but to a capacity the employment of which *precedes* adopting the intentional stance.

Adopting the intentional stance, according to Dennett (2006), is a capacity available to higher animals. These do not only detect agents, but can also discriminate between sorts of behaviour: will it attack or flee, will it back down when I threaten it, does it want to eat me or my neighbour? Dennett: "These clever animals have discovered the *further* [my emphasis] Good Trick of *adopting the intentional stance* [Dennett's emphasis]" (2006, p. 109).

The position that emerges from a combination of IST and ER differs in a number of crucial respects from Dennett's current position. Let me fill in some details about the combined IST + ER position by highlighting two important differences.

To start with, the function Dennett assigns to ADD's differs subtly from the function I assign to ER. ADD's serve merely to indicate whether a system is an intentional system. All further questions about the nature of the perceived intentionality of behaviour are to be answered by adopting the intentional stance. ER, by contrast, does not merely indicate that a system displays intentional behaviour, but also allows us some insight into the nature of the perceived intentional behaviour. Through empathic resonance we do not just perceive 'pure intentionality'. Rather we perceive, e.g. anger, fright, shyness, a cooperative attitude, an intention to shake hands or open a door, etc. This is a good reason to favour ER over ADD's—in Dennett's sense—as a necessary step preceding the use of an intentional stance: it seems highly artificial to claim that animals are capable of perceiving intentionality in behaviour without being aware at all of the *kind* of intentionality involved.

The second difference is even more important from the perspective of the current project. Through empathic resonance, we have indirect access to the urges or drives behind the behaviour of others. I want to insist on the idea that the agents whose behaviour is observed via ER usually *experience* these urges or drives. There usually is a phenomenal aspect to anger, fright, shyness and other motivations accessible via ER. Often, but certainly not always, this phenomenal aspect is 'mirrored' in the empathically resonating observer (see the quote by Gordon above). But here I am concerned with the phenomenal character of the motivating drive or urge behind the behaviour of the observed person.

The reason that I insist on the phenomenal character of these motivations is the implication that the internal states of others tracked via ER are at least in some sense *mental*. This will be crucial when it comes to the question of mental causation.

Before I will discuss this, let me emphasize that this particular aspect of the IST + ER proposal is incompatible with Dennett's oeuvre. When it comes to the phenomenal character of mental states, Dennett basically is an eliminativist (Dennett 1988, 1991b; Dennett and Kinsbourne 1992). This, however, does not make my proposal incoherent: although Dennett's position on phenomenal consciousness does presuppose IST, the reverse is not true. IST is completely independent of Dennett's views on consciousness.

4 Phenomenal Causation

The claim I wish to make in this brief section is that the position described above allows at least for the possibility of a form of mental causation, sometimes called phenomenal causation (Tye 1995). The main observation that is required to make this claim is that the ontological status of the 'urges' and 'drives' behind behaviour to which we have indirect access through ER—the motivations that are discerned *before* adopting the intentional stance in stereotypical situations—is fundamentally different from the ontological status of the full-blown propositional attitudes we attribute using the intentional stance. While propositional attitudes are, according to IST, useful *interpretations* of behaviour, a *heuristic overlay*, these subjectively felt urges and drives are real internal states of the agent. Interpretations and heuristic overlays cannot cause anything, but real internal states *can*.

Let me be clear about what the proposal is here. The claim is *not* that ER is involved in mental causation. Rather, the claim is that by making ER compatible with and required by IST—as I have tried to do above—the combined IST + ER theory *implies* the reality of phenomenal motivations behind intentional action in a way that suggests phenomenal causation. It is quite possible that the subjective 'feels' behind intentional actions are actually causally involved in the production of those actions. It certainly is possible that the anger I feel is causally responsible for my shouting, that the positive feeling I have when recognizing my children at the schoolyard causes me to smile, etc. These would be instances of phenomenal causation.

The type of mental causation that I am suggesting might be involved here does *not* exactly match the Darwin–James-like view of intention-in-action. For in that view the basic intention more or less coincides with the action. Intuitively at least, a time lag between intention and action is required for causation. But of course, perceiving temporally separate occasions of basic intentions-in-action does allow for the required time lag and hence may justify using the Darwin–James-like view in connection with phenomenal causation. For instance, I may perceive anger in someone's facial expression and shouting by that person a few seconds later, in which case I may infer that the shouting may have been caused by the anger.

Of course the actual causal efficacy of phenomenal urges and drives is not something that I can argue for here. It is certainly possible that such a phenomenal feel is itself the effect of the actual brain state that is causally efficacious in producing the action associated with the feel (see, e.g., Kim 1998, pp. 70–72). Psychologists such as Daniel Wegner claim that such epiphenomenalism is the normal situation (Wegner 2002; NB: Dennett appears to be largely in agreement,

2003b), but this is controversial (see e.g. Nahmias 2002; Bayne 2006). This is not the place to try and settle that controversy, however. My claim in this section is that the possibility that these urges and drives cause behaviour is in no way contradicted by IST + ER. Therefore, an IST-like position that includes a form of mental causation is an option.

5 The Causal Relevance of Propositional Attitudes

Phenomenal causation is a form of mental causation. But it is *not* the kind that is the prime concern of the majority of philosophers active in the debate on mental causation. The kind of causation that occupies centre stage in that debate is causation of actions by propositional attitudes, e.g. beliefs and desires.

Admittedly, according to the theory of the mental that results from combining IST with ER, propositional attitudes have no causal efficacy. Attributions of beliefs and desires, according to that theory, are *interpretations*—interpretations of unconceptualized internal motivational states accessed via ER. And even though they are interpretations of internal states that may themselves be causally efficacious in producing actions, interpretations are as such not directly causally efficacious. The purpose of this section, however, is to claim that though propositional attitudes are not efficacious in the proposal under discussion, they *are* what Jackson and Pettit would call *causally relevant*. Whether that is sufficient to satisfy our intuitions about mental causation will be discussed in the next section.

The notion of causal relevance was introduced in the context of Jackson and Pettit's 'program model' (Jackson and Pettit 1988, 1990; Pettit 1993), an attempt to grant multiply realisable higher order properties such as states defined in terms of their causal roles (e.g., according to functionalism, mental states) a job in causal explanations, while acknowledging that the real causal work in this world is being done in the micro-physical realm. For instance: even though all causal efficacy is in that realm, the elasticity of an eraser is causally relevant to it's bending (Pettit 1993, p. 33). The basic idea is the following: correctly attributing to an entity or system a state defined in terms of a causal role (e.g. attributing elasticity to an eraser, or, assuming functionalism, attributing a belief to a person) secures the presence of *some* realizer that plays the causal role in terms of which the attributed state is defined. Elasticity, being a supervenient property, may not be a causally efficacious property, but attributing it to an eraser means attributing a micro-physical structure that plays the causal role of being elastic. In Jackson and Pettit's terms: elasticity programs for certain 'behaviour' under certain circumstances. Which is why it can function in causal explanations without itself being causally efficacious.

The notion of causal relevance was advanced in the context of supervenience relations allowing higher order properties to be multiplying realisable by various types of configuration of micro-physical particles. In view of Kim's attack on multiple realisation (1992) and his introduction of the orders/levels distinction (1998), it might seem problematic nowadays. The use I will make of it here is not based on multiply realisable supervenience relations, but on relations of interpretation.

I believe that the notion of causal relevance also applies to propositional attitudes in the theory that combines IST with ER. The idea here is that the appropriateness or the *fruitfulness* of an interpretation of an internal motivational state in terms of, say, a desire *D* and belief *B*, signals the presence of an internal state with a specific causal profile approximated by *D* and *B*. Of course the term 'approximation' makes for a significant difference with the program model in which *D* and *B* would be completely captured by their supervenience base, thus securing a nomological connection between *D* and *B* and the causally efficacious base. Such a nomological connection is not to be had in the case of mere approximation. Nevertheless, I believe the connection between *D* and *B* on the one hand and the phenomenal motivations they signal on the other, to be reliable enough to warrant the term 'causal relevance'. Folk-psychology is immensely successful in capturing the causal profiles of our motivations. The fact that it is not infallible only underlines, in my opinion, that IST + ER yields a more realistic view of it than does the program model.

Thus, like with the program model, when the state a person is in is best described in terms of a (set of) propositional attitude(s), that reliably indicates the presence of some internal state that will cause the actions that *D* and *B* help to predict. In the combined IST + ER theory, propositional attitudes are causally relevant despite being causally inefficacious.

6 Is this Sufficient to Satisfy our Intuitions about Mental Causation?

A program model-like view on the mental causation issue would be compatible with Dennett's IST too, to be sure. The difference between IST and IST + ER in this respect, is 'merely' the fact that on the latter theory the internal, causally efficacious states are considered unconceptualized motivational (i.e. mental) states, due to their specific phenomenal character. On IST, by contrast, they are mere brain states.

This difference is crucial, I will argue in this last section. Even *if* Dennett would embrace the idea of causal relevance, when it comes to accounting for intuitions about mental causation that would be his only resource. He would have to claim that causal relevance explains *all* our intuitions about mental causation. I will argue that this is unrealistic by showing how IST + *ER can* account for these intuitions only by assigning important work to the phenomenal character of the internal motivational states that are being interpreted using folk-psychology. Thus, the explanatory burden on the shoulders of 'merely' causally relevant beliefs and desires becomes more limited and realistic.

I take an adequate account of what is at stake in the debate on mental causation to explain at least three things: (1) Most of our actions can be rationalized using various reasons. Usually, though, there is only one reason for which we did an action. According to Davidson (1963) and a majority of analytical philosophers, the way to single out *the* reason for which we did an action is to claim that this reason was the one that caused our action. (2) Practical reasoning, the outcome of which is the formation of intentions, matters to the way we act. According to most, this is because intentions, e.g. belief-desire pairs, cause actions. (3) There is a phenomenology of *doing*. This occupies centre-stage in Daniel Wegner's (2002) book

against the reality of what he calls 'conscious will' (a notion that overlaps with mental causation rather than with free will). The intuition here is that while acting intentionally and consciously, there is an *experience* (rather than belief) of 'doing', an experience of the actions done as *mine*. Wegner: "Consciously willing an action requires a feeling of doing (…), a kind of internal "oomph" that somehow certifies authentically that one has done the action" (Wegner 2002, p. 4).[1]

The first two explananda are about reasons for action, i.e. about propositional attitudes explaining actions and not so much about the phenomenology of agency. The third explanandum, by contrast, is not about reasons or propositional attitudes, but about an 'oomph', an *experience* of being an agent, the 'author' of one's actions.

Starting with the third explanandum, I believe there is good reason to relieve propositional attitudes of a potential explanatory task here. Recent research into the phenomenology of intentional agency highlights the point that first experiencing oneself to form an intention at the level of propositional attitudes (say, on the basis of a belief and a desire) and consequently experiencing oneself to act as intended is actually very rare. Moreover it is not an experience of *doing* (Horgan et al. 2003; Horgan and Tienson 2002, 2003; Bayne 2006; Bayne and Levy 2006). What we experience is the conjunction—in Humean terms—of an intention and an action. Perhaps, if we experience the conjunction as *causation*, it would be an experience of doing.[2] But even on a Humean theory of causation in terms of *constant* conjunction, experiencing conjunctions of intentions and actions is merely enough to inductively *infer* mental causation, given that *constant* conjunction cannot be experienced. Explaining the experience of doing cannot be done in terms of actions matching prior intentions. Hence, whether intentions, or beliefs and desires, that precede actions are causally efficacious or causally relevant, is irrelevant in this specific context.

Can the third explanandum, Wegner's 'oomph', be explained in terms of the Darwin-James-like view of intention-in-action? I believe it can in a straightforward fashion: It is very intuitive to hold that Wegner's 'oomph' and the unconceptualized phenomenal motivational states we are able to 'pick up' via ER—the urges and drives—are different ways of referring to the same kind of experience. What we pick up on when we empathically resonate with other people, I submit, are precisely their Wegnerian 'oomphs'.

I now turn to the first two explananda. Given that these are about propositional attitudes there is no point trying to explain them in terms of the phenomenal causation of Sect. 4. But they *can* be explained in terms of the causal relevance of propositional attitudes.

[1] As mentioned in Sect. 4, Wegner thinks it is an illusion that this 'oomph' indicates real causal efficacy. Others (e.g. Bayne 2006) think it is at least possible that there is no illusion involved here. It is the compatibility of that last option with IST (by adding ER) that I am arguing for here.

[2] In arguing against a Humean conception of causation, Thomas Reid exploited this observation. Causation, he insisted, must not just be conceived of in terms of a conjunction of separated events; there must be some inner connection too. Reid noted that possibly the only instance in which such a connection can be experienced, is in the experience of doing. Thus, insofar as the concept derives from experience, "the conception of an efficient cause may very probably be derived from the experience we had (…) of our own power to produce certain effects (…)" (Reid 1785, quoted by Chisholm 1982, p. 31).

Causal relevance of propositional attitudes can deal with the first explanandum in the following way: There are good and bad folk-psychological interpretations of unconceptualized internal motivational states. The best interpretation is the one that predicts a range of actions that is closest to the factual range of actions the internal states causally allows for. The idea here simply is that the best interpretation picks out *the* reason for which a person did an action from the range of possible reasons. This does not require any causal efficacy on the part of the attributed beliefs and desires. But obviously, there should be causal *relevance* to the attributed beliefs and desires.

Indeterminacy of reason attribution can be considered a problem here. IST allows for the possibility of rivalling reason attributions that are equally adequate when it comes to action prediction. In cases of indeterminacy, on IST, there is not *one* reason for which a person did something. The problem is not serious: It must be emphasized that even Dennett thinks such cases are extremely rare or even just theoretical.

In order to see that causal relevance of propositional attitudes can deal with the second explanandum, we must be aware of the nature of practical reasoning according to IST + ER. Practical reasoning is making inferences involving propositional attitudes. On IST + ER, entertaining propositional attitudes is in fact interpreting ones own urges and drives. We can think of a process of practical reasoning in terms of beliefs and desires interacting. But the view of the interpretationist position implied here, is that practical reasoning involves focussing on some urges, disregarding others, even letting new kinds of urges emerge and finding a strategy of action that does justice to most or the most urgent (literally) ones (we need not think of this in terms of a homunculus who is busy selecting some urges and blocking others, of course). This process can be *described* as beliefs and desires interacting, but also as a form of self-interpretation. The outcome of such a process can be portrayed as the formation of an intention to act or as the desire for x and the belief that y-ing will achieve x.

The point is that considering this intention or belief/desire pair or preceding beliefs and desires to be merely causally relevant in no way implies the rejection of the fact that some urges are focussed on and others ignored; practical reasoning does matter to how we act. But once again, causal relevance is required for us to be able to conceive of practical reasoning in folk-psychological terms: e.g. for the outcome of a process of practical reasoning to be describable as an intention, the causal role definition of the intention must actually match the causal profile of the relevant internal state.

The one question that remains is whether there is something missing from the above account of the explananda involved in the debate over mental causation, due to the fact that propositional attitudes are merely causally relevant. My answer is negative.

The intuition that something *is* missing might be phrased as follows. Most of us identify with our minds rather than with our bodies (Bloom 2004; people usually say they *are* their minds and *have* their bodies). And when thinking about minds we usually think about our thoughts; the intuition is "I am my thoughts". Hence there is a tendency to conclude that "when *I* do something, my thoughts must be causally efficacious in producing an action." Now of course the concept 'thoughts' is vague

here. A natural interpretation of it is in terms of beliefs, desires and other propositional attitudes. Under *that* interpretation, what happens is that the third explanandum needs to be being taken care of at the level of full blown propositional attitudes. As we have seen, that is not possible.

The real intuitive problem here is not that causally relevant propositional attitudes are insufficient for explananda 1 and 2, it is that they are insufficient for explanandum 3. This may be a problem for IST, but not for IST + ER. For on that proposal, explanandum 3 is being explained in terms of the Darwin–James-like view on intention-in-action. So, it is because explanandum 3 is taken care of in a phenomenologically adequate fashion by the ER-component of the theory under discussion that the explanation of explananda 1 and 2 in terms of causal relevance by the IST-component suffices.

Acknowledgement I would like to thank an anonymous referee for this journal for very helpful comments.

References

Baron Cohen, S. (1995). *Mindblindness*. Cambridge, MA: MIT Press/Bradford Books.

Bayne, T. (2006). Phenomenology and the feeling of doing: Wegner on the conscious will. In S. Pockett, W. P. Banks, & S. Gallagher (Eds.), *Does consciousness cause behavior? An investigation of the nature of volition* (pp. 169–186). Cambridge, MA: MIT Press.

Bayne, T., & Levy, N. (2006). The feeling of doing: Deconstructing the phenomenology of agency. In N. Sebanz & W. Prinz (Eds.), *Disorders of volition*. Cambridge, MA: MIT Press.

Bloom, P. (2004). *Descartes' baby: How the science of child development explains what makes us human*. London: William Heinemann.

Carr, L., Iacoboni, M., Dubeau, M. C., Mazziotta, J. C., & Lenzi, G. L. (2003). Neural mechanisms of empathy in humans: A relay from neural systems for imitation to limbic areas. *Proceedings of the National Academy of the Sciences, 100*, 5497–5502.

Chisholm, R. (1982). Human freedom and the self. In G. Watson (Ed.), *Free will*. Oxford: Oxford University Press.

Damasio, A. R. (1994). *Descartes' error: Emotion, reason and the human brain*. New York: Putnam's Sons.

Darwin, C. (1872/1965). *The expression of emotion in man and animals*. Chicago: Chicago University Press.

Davidson, D. (1963). Actions, reasons, and causes. *Journal of Philosophy, 60*, 685–699.

Dennett, D. C. (1969). *Content and consciousness*. London: Routledge.

Dennett, D. C. (1978). *Brainstorms*. Hassocks: Harvester Press.

Dennett, D. C. (1987). *The intentional stance*. Cambridge, MA: MIT Press.

Dennett, D. C. (1988). Quining Qualia. In Marcel, A. & Bisiach, E. (Eds.), *Consciousness in contemporary science*. Oxford: Oxford University Press.

Dennett, D. C. (1991a). Real patterns. *The Journal of Philosophy, 88*, 27–51.

Dennett, D. C. (1991b). *Consciousness explained*. Boston: Little Brown.

Dennett, D. C. (2003a). *Freedom evolves*. London: Allen Lane.

Dennett, D. C. (2003b). Making ourselves at home in our machines: The illusion of conscious will. *Journal of Mathematical Psychology, 47*, 101–104.

Dennett, D. C. (2006). *Breaking the spell*. New York: Viking.

Dennett, D. C., & Kinsbourne, M. (1992). Escape from the Cartesian theater. *Behavioral and Brain Sciences, 15*, 183–247.

Fadiga, L., Fogassi, L., Pavesi, G., & Rizzolatti, G. (1995). Motor facilitation during action observation: A magnetic simulation study. *Experimental Brain Research, 73*, 2608–2611.

Fodor, J. (1985). Fodor's guide to mental representations. *Mind, 94*, 76–100.
Gallagher, S. (2004). Understanding interpersonal problems in autism: Interaction theory as an alternative to theory of mind. *Philosophy, Psychiatry, & Psychology, 11*, 199–217.
Gallagher, S. (2005). *How the body shapes the mind*. New York: Oxford University Press.
Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain, 119*, 593–609.
Gallese, V., & Goldman, A. (1998). Mirror neurons and the simulation theory of mind reading. *Trends in Cognitive Sciences, 2*, 493–501.
Goldman, A. I. (2006). *Simulating minds: The philosophy, psychology and neuroscience of mindreading*. New York: Oxford University Press.
Gordon, R. M. (1996). 'Radical' simulationism. In P. Carruthers & P. K. Smith (Eds.), *Theories of theories of mind* (pp. 11–21). Cambridge: Cambridge University Press.
Harris, P. L. (1991). The work of the imagination. In A. Withen (Ed.), *Natural theories of mind: The evolution, development and simulation of everyday mindreading*. Oxford: Basil Blackwell.
Harris, P. L. (1992). From simulation to folk-psychology: The case for development. *Mind and Language, 7*, 120–144.
Hatfield, E., Cacioppo, J. T., & Rapson, R. L. (1994). *Emotional contagion*. Cambridge: Cambridge University Press.
Heal, J. (1986). Replication and functionalism. In J. Butterfield (Ed.), *Language, mind, and logic* (pp. 135–150). Cambridge: Cambridge University Press.
Heal, J. (1996). Simulation, theory, and content. In Carruthers, P. & P. K. Smith (Eds.), *Theories of theories of mind* (pp. 75–89). Cambridge: Cambridge University Press.
Horgan, T., & Tienson, J. (2002). The intentionality of phenomenology and the phenomenology of intentionality. In D. Chalmers (Ed.), *Philosophy of mind: Classical and contemporary readings* (pp. 520–533). New York: Oxford University Press.
Horgan, T., & Tienson, J. (2003). The phenomenology of embodied agency. In J. Joao Saagua (Ed.), *The explanation of human interpretation* (pp. 415–424). Lisbon: Colibri (in Portuguese).
Horgan, T., Tienson, J., & Graham, G. (2003). The phenomenology of first-person agency. In S. Walter & H. D. Heckmann (Eds.), *Physicalism and mental causation: The metaphysics of mind and action* (pp. 323–340). Charlottesville: Imprint Academic.
Iacoboni, M. (2003). Understanding others: Imitation, language, empathy. In Hurley, S. & Chater, N. (Eds.), *Perspectives on imitation: From cognitive neuroscience to social science: Mechanisms of imitation and imitation in animals*, (Vol. 1). Cambridge, MA: MIT Press.
Jackson, F., & Pettit, P. (1988). Functionalism and Broad Content. *Mind, 97*, 381–400.
Jackson, F., & Pettit, P. (1990). Program explanation: A general perspective. *Analysis, 50*, 107–117.
James, W. (1890/1984). What is an emotion? In C. Calhoun & R. C. Solomon (Eds.), *What is an emotion?* (pp. 125–142). New York: Oxford University Press.
Kim, J. (1992). Multiple realizability and the metaphysics of reduction. *Philosophy and Phenomenological Research, 52*, 1–26.
Kim, J. (1998). *Mind in a physical world: An essay on the mind–body problem and mental causation*. Cambridge, MA: MIT Press.
McCulloch, G. (1990). Dennett's little grains of salt. *The Philosophical Quarterly, 40*, 1–12.
Melzoff, A. N., & Gopnik, A. (1993). The role of imitation in understanding persons, developing a theory of mind. In S. Baron-Cohen, H. Tager-Flusberg, & D. J. Cohen (Eds.), *Understanding other minds: Perspectives from autism*. Oxford: Oxford University Press.
Melzoff, A. N., & Moore, M. K. (1977). Imitation of facial and manual gestures by human neonates. *Science, 198*, 75–78.
Nahmias, E. (2002). When consciousness matters: A critical review of Daniel Wegner's the illusion of conscious will. *Philosophical-Psychology, 15*, 527–541.
Pellegrino, G. D., Fadiga, L., Fogassi, L., Gallese, V., & Rizzolatti, G. (1992). Understanding motor events. *Experimental Brain Research, 91*, 176–180.
Perner, J. (1996). Simulation as explanation of predication-implicit knowledge about the mind: Arguments for a simulation-theory mix. In P. Carruthers & P. K. Smith (Eds.), *Theories of theories of mind* (pp. 90–104). Cambridge: Cambridge University Press.
Pettit, P. (1993). *The common mind*. New York: Oxford University Press.
Provine, R. R. (1986). Yawning as a stereotypical action pattern and releasing stimulus. *Ethology, 72*, 109–122.

Provine, R. R. (1989). Contagious yawning and infant imitation. *Bulletin of the Psychonomic Society, 27*, 125–126.

Provine, R. R. (1992). Contagious laughter: Laughter is a sufficient stimulus for laughs and smiles. *Bulletin of the Psychonomic Society, 30*, 1–4.

Reid, T. (1785). *Essays on the intellectual powers of man*. Edinburgh/London: John Bell/G.G.J. & J. Robinson.

Rizzolatti, G., Fadiga, L., Gallese, V., & Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Cognitive Brain Research, 3*, 131–141.

Searle, J. R. (2001). *Rationality in action*. Cambridge, MA: MIT Press.

Shoemaker, S. (1980). Causality and properties. In P. Van Inwagen (Ed.), *Time and cause: Essays presented to Richard Taylor* (pp. 109–136). Dordrecht: Reidel.

Tye, M. (1995). *Ten problems of consciousness*. Cambridge, MA: MIT Press.

Wegner, D. (2002). *The illusion of conscious will*. Cambridge, MA: MIT Press/Bradford Books.