

A problem for Wegner and colleagues' model of the sense of agency

Glenn Carruthers

Published online: 17 February 2010
© Springer Science+Business Media B.V. 2010

Abstract The sense of agency, that is the sense that one is the agent of one's bodily actions, is one component of our self-consciousness. Recently, Wegner and colleagues have developed a model of the causal history of this sense. Their model takes it that the sense of agency is elicited for an action when one infers that one or other of one's mental states caused that action. In their terms, the sense of agency is elicited by the inference to apparent mental state causation. Here, I argue that this model is inconsistent with data from developmental psychology that suggests children can identify the agent behind an action without being capable of understanding the relationship between their intentions and actions. Furthermore, I argue that this model is inconsistent with the preserved sense of agency in autism. In general, the problem is that there are cases where subjects can experience themselves as the agent behind their actions despite lacking the resources to make the inference to apparent mental state causation.

Keywords Sense of agency · Self-consciousness · Wegner · Sense of self

Introduction

In this paper, I examine a model of the sense of agency over one's bodily actions that is presented by Wegner and colleagues. This model provides an account of the causal history of the sense of agency, i.e. it explains how the sense is elicited. Unlike some alternative models, such as those presented by Frith et al. (2000) or Synofzik et al. (2008), it does not tie this directly to motor control, but rather to the ability to draw inferences about one's mental states as the cause of one's actions. This could lead to some confusion as to what is being explained. To some readers, it may appear that the target of explanation is these inferences or, at best, judgements of agency and not the sense of agency. In light of this, I begin with some textual evidence

G. Carruthers (✉)
Macquarie Centre for Cognitive Science (MACCS), Macquarie University, Sydney 2109, Australia
e-mail: gcarruth@maccs.mq.edu.au

which suggests that these authors do intend to provide an account of the sense of agency. Following this, I present their model and the data used to motivate and support it. Having presented what I take to be the strongest case for this model, I turn to the main problem with it. I present a number of cases which suggest that young children and some of those suffering from autism have an intact sense of agency despite not being able to infer that their mental states cause their actions. This dissociation suggests that it is not this inference that causes the sense of agency to be elicited. Given that the model presented by Wegner and colleagues seems unable to explain every instance of the sense of agency, the possibility of an alternative explanation of the data it can explain gives us good reason to look elsewhere for an explanation.

Inference to apparent mental state causation and the sense of agency

Wegner and colleagues offer a model of how the sense of agency is elicited. They take it that the sense of agency arises from an inference to the cause of an action. If one of one's own mental states (often one's intention, but not always) is inferred to be the cause of an action, then, one experiences a sense of agency over that action (Wegner 2003, p. 67). However, this may be a controversial reading of the model. It may be possible to read this model as an account of how we come to judge that one or other of our mental states causes our actions, rather than the sense of agency. This interpretation is suggested by all the work judgements of "apparent mental causation" do in the theory (Wegner 2003, p. 65). In this section, I clarify Wegner and colleagues' explanatory target and show how an account of the sense of agency is intended to be a part of their model. I also examine Wegner and colleagues' conception of the relationship between the sense of agency and judgements of agency.

At times the explanatory target of Wegner and colleagues' model does seem to be judgements of apparent mental causation alone. For example:

Every few moments of every day, we think about doing something and then do it. We think of moving a finger and then do it, we think of going to the store for milk and do it, we think of looking away from this page- and then do it. It certainly doesn't take a rocket scientist to draw the obvious conclusion from a lifelong accumulation of such examples: consciousness is an active force, an engine of will (Wegner 2003, p. 65)

... people interpret their own accessible thought as the cause of the behavioural event, thereby arriving at the deceptive belief that they produced it (Aarts et al. 2005, p. 442).

Here, it seems that Wegner and colleagues are discussing the belief that conscious thoughts cause action in virtue of being those thoughts and not the sense of oneself as being the agent of the action.

At other times, Wegner and colleagues seem to set their sights on judgements about who is the agent of an action:

Most people can readily sort many events in the world into those they have authored and those they have not. This observation suggests that each person

has a system for authorship processing... a set of mental processes that monitors indications of authorship to judge whether an event, action, or thought should be ascribed to self as a causal agent (Wegner et al. 2004 p. 838 emphasis mine).

Similarly, Wegner and Wheatley (1999, p. 487) make use of Spanos' (1982) description of subjects' understanding of action:

Interpreting behaviour as an action involves attributing causality to the self (e.g. I did it), while interpreting it as a happening requires that causality be attributed to sources other than the self (e.g. It happened to me) (Spanos 1982, p. 200).

From these quotes, it seems that the explanatory target of Wegner and colleagues' model is not the sense of agency, but rather two judgements, these being the judgement to apparent mental state causation and judgements of agency.

However, this is not the complete picture of their description of the phenomena. For example, when referring to Penfield's famous study on eliciting movement in subjects by direct stimulation of the brain, Wegner describes the subjects' reports as follows:

...their common report of the experience was that they did not 'do' the action, and instead felt that Penfield had 'pulled it out' of them (Wegner 2003, p. 65).

Here, the subjects' experience seems to be that they did not do the act, as well as the experience of not having an intention to do the act. This seems to involve the subjects' experience of or judgements about their own agency, not just what mental state caused the action. But, is this to be considered a feeling like the sense of agency or an attribution judgement? Wegner and colleagues seem to opt for the prior option, describing the phenomenon as a feeling (Wegner 2002, pp. 3, 29) or, more commonly, as the "authorship emotion" (e.g. Wegner 2002, p. 326; Wegner et al. 2004, p. 840):

... when we sense our own authorship of an action, there is an additional quality present, a feeling of doing, that marks the event uniquely. The feeling of consciously willing adds a psychological exclamation point ("I ran the stoplight!").

The experience of consciously willing an action has been described as an authorship emotion (Wegner 2002), a feeling that ties the basic fact of the causal event to a bodily response... The experience of conscious will need not be a veridical expression of how the action came about (although people tend to interpret it this way), but it does serve to authenticate the action as something done by the self. Authorship seems to be a self-recognition of agency, then, that has both a rational component (knowledge that one was the agent causing the action) and an experiential component (the feeling of consciously willing the action) (Wegner et al. 2004, p. 839 my emphasis).

Here is a clear statement that the sense of agency is part of the phenomena of interest. The sense of agency is here being described as the "experiential component" of authorship or the "authorship emotion" as evidenced by the fact

that this experience is described as something which is used to “authenticate the action as something done by the self” (Wegner et al. 2004, p. 839). This is precisely the role of the sense of agency. Indeed, Wegner and colleagues’ distinction between a “rational” and “experiential” component of authorship seems to pre-empt Synofzik et al. (2008) distinction between the judgement of agency and the feeling (sense) of agency. However, we may wonder what Wegner and colleagues’ model aims to explain. Is it the rational or the experiential component of authorship? Immediately after the above passage, Wegner and colleagues remove such doubt, stating:

The development of the experiential component of one’s own action authorship is the focus of the theory of apparent mental causation. According to this theory, the feeling that we consciously will our own actions is traceable to an inference we make from the match between our conscious thoughts and observed action (Wegner et al. 2004, p. 839 my emphasis).

Here, Wegner and colleagues state, in no uncertain terms, that it is the experiential component of authorship, i.e. the sense of agency, which is to be explained. Furthermore, they tell us how it is to be explained. For Wegner and colleagues, the sense of agency is to be explained by judgements one makes about one’s mental states causing one’s actions.

So, it seems that the general strategy Wegner and colleagues adopt is to explain the sense of agency via judgements of apparent mental state causation; but, what about judgements of agency? Where do they fit in this picture? To see, we can compare this explanation to two other strategies for explaining the sense of agency. On the first alternative, the sense of agency is elicited by the motor control system. This is the case in comparator model discussed below. On this kind of model, the sense of agency is elicited by the motor control system prior to judgements of agency being formed. The second alternative is subtly different to Wegner and colleagues approach, and it is not easy to distinguish the two. This approach, exemplified by Stephens and Graham (2000) account of the sense of agency over thought, explains the sense of agency in terms of a judgement or inference, but it is not a judgement to apparent mental state causation, rather it is a judgement to whom the agent of action is. On this second approach, a judgement of agency elicits the sense of agency.

Although at first glance, the view of the relationship between judgements and the sense of agency advocated by Wegner and colleagues seems more in line with the second of these options, it is not. The reason for this is that the content of the two judgements is different. On the Wegner and colleagues approach, the judgement is about a particular mental state causing an action, whereas on the Stephens and Graham type view, the judgement is about who the agent is. The agent is already in the content of the judgement for the later conception of the relationship. However, on Wegner and colleagues’ conception, an additional step is needed, that being the elicitation of the sense of agency.

What about the quotes that seem to suggest that it is judgements of agency rather than the sense of agency which is the explanandum of their theory? These could be simple slips of the pen, a misaligning of the sense of agency and judgements used to operationalise this sense (as done in the experiments discussed below). Otherwise, they could be an additional theoretical claim. However, this seems unlikely as it

would involve the addition of a judgement of agency between the judgement of apparent mental state causation and the sense of agency. What is clear in text quoted above is the proposal as to the relationship between judgements of apparent mental state causation and the sense of agency, namely the judgement of apparent mental state causation elicits the sense of agency. Is this a viable account of how the sense of agency is elicited? Let us examine the account in some detail.

Wegner and colleagues' model

Wegner and colleagues suggest that whenever one infers that one's intention is the cause of an action one experiences a sense of agency over that action (Wegner 2003, p. 67; Wegner and Wheatley 1999, p. 480). If I experience my intention to get a coke, and then I go and get one, I can infer that my intention caused me to get one. These inferences to "mental state causation" are not themselves the sense of agency, rather, Wegner and colleagues seem to be claiming that the sense of agency arises out of or is elicited by such inferences.

For Wegner and colleagues, if I am to feel that I am the one writing this sentence, I must first infer that my intention to write this sentence caused this sentence to be written. According to these authors, one makes this inference if three conditions are met. First, the mental state must seem to be consistent with the action. My intention to get a drink has nothing to do with my sitting down to work and thus could not be inferred to be a cause of that action. However, my intention to sit down to work could be. Call this the principle of consistency. Second, the mental state must seem to occur before the action. My desire to get a drink now has nothing to do with my getting the drink I just finished, so I cannot be inferred as the cause of that action. However, my desire to get a drink that occurred just before I got that drink could be. Call this the principle of priority. Third, the mental state must be the only apparent cause of the action. My desire to sit down would not be inferred as the cause of my sitting if someone had pushed me into the chair. Call this the principle of exclusivity (Wegner 2003, p. 67; Wegner and Wheatley 1999, pp. 482-487).

Some support for the principle of exclusivity can be found when we examine instances of how the sense of agency changes. Imagine that you are moving swiftly down a flight of stairs. At the appropriate floor, you slow and reach for the door handle that will allow you egress from the stairwell. As you open the door, you find it moves far more rapidly than you had intended. Someone else is opening the door from the other side! Call this the simultaneous door-opening effect. This effect involves the feeling that one is the agent behind an action being suddenly replaced by the feeling that one is not the agent due to the interference of another. According to Wegner and colleagues' model, this effect occurs due to a perceived shift in the relationship between one's intentions and the event. When one first reaches for the door, one's intention to do so is the only plausible cause of that action. The principles of exclusivity, priority, and consistency are met. However, when one becomes aware of someone else opening the door from the other side this condition is no longer met. The other person appears to be as much as cause of the door opening as one's intention. The exclusivity principle is not met, and thus, one cannot infer that one's intention caused the door to open; as a consequence, the sense of being the agent behind the door opening is lost (Wegner 2002, pp. 90-95).

This model can also provide an explanation of the apparently mystical phenomenon of Ouija board spelling. The use of the Ouija board involves two or more participants moving a pointer over a board with the letters of the alphabet, numbers, and the words “yes” and “no” printed. The participants focus on a question. Often, it is found that a coherent response is spelt out by the board. This is usually attributed to some other-worldly source—often a deceased relative (Wegner 2002, pp. 108–112). Explanations that take this phenomenon seriously (that is, explanations that do not assume all instances of Ouija board spelling involve deception) take it that the participants move the pointer without experiencing themselves as the agent(s) behind the movement. Wegner and colleagues’ model provides an explanation of why the sense of agency is absent under these conditions. Wegner notes that users of the Ouija board usually begin a session by moving the pointer in a random pattern or a basic shape (such as a circle; Wegner 2002, p. 110). As the movement continues, each participant attempts different movements to the others, resulting in a movement that is not intended by any one person involved. The principle of consistency is violated, and thus, each participant has reduced grounds to infer that their mental states caused the movement; thus, they do not experience themselves as the agent behind the movement. Furthermore, the exclusivity principle is not met. Each of the participants is not an exclusive cause of the movement. All (or at least some) of the participants are involved in moving the pointer. As such, no one participant could sensibly infer that they are the sole cause of the movement. Furthermore, those who report this phenomenon commonly claim to adamantly believe that the Ouija board provides a way to communicate with spirits. This belief alone is enough to violate the exclusivity principle as it provides another potential source of the movement. As the principles of exclusivity and consistency are not met in the case of Ouija board spelling, Wegner and colleagues’ model predicts that participants should not feel a sense of agency over the movements of the pointer (Wegner 2002, p. 112). The ability to explain the conditions under which feelings of agency apparently change offers considerable support to Wegner and colleagues’ model.

Wegner and colleagues have expanded on this reasoning across a number of studies. The “helping hands” experiment (Wegner et al. 2004) is an illustrative example. In this study, one subject (the participant) stood with their back against the second subject (the helper). The participant stood with their arms by their side, whilst the helper reached their arms forward underneath the participant’s arms. A screen obscured the helper in such a way that their arms appeared (from the front) to be those of the participant. Both subjects were given headphones. The helpers heard a series of instructions to perform a set of hand movements (e.g. make the ok sign with both hands). The participants were divided into three groups depending on what they heard. One group (the preview group) heard the instructions to the helper whilst the other groups (the control groups) heard either nothing or an instruction to perform a movement other than what the helper heard. Those in the preview group reported a greater sense of agency over the movements of the helper than either of the control groups (Wegner et al. 2004, pp. 841 & 842).

The inference to apparent mental state causation account of the sense of agency can explain these findings. Those in the preview group mistakenly infer that the experience of hearing the instruction (or perhaps a motor representation of the action

caused by hearing the instruction)¹ causes the action to take place. This occurs because the instruction is prior to and consistent with the seen action. It is consistent in that it represents the same action. This mistaken inference causes the subjects to feel that they are involved in generating the action. However, they do not experience themselves as the sole agent of the action as they know the helper is the one moving. As such, the principle of exclusivity is not met. In contrast, those in the control groups do not infer that any thought causes the observed action as they did not hear an instruction or, if they did, it was an instruction to perform a different action to the one they see. As such, the principle of consistency is not met.

Along similar lines to this study is the task known as 'I-spy'. As in the helping hands task, in the I-spy task Wegner and Wheatley (1999) attempt to elicit a sense of agency for actions the subject does not control. For this task subjects sat next to a confederate (whom they believed was another participant) looking at the same computer screen. On the screen was a picture of various toys (a dinosaur, a swan and the like). The subject and the confederate held a small square board over a typical computer mouse. They were asked to use this together to move a cursor in slow circles around the images on the screen. On each trial the subject and confederate would move the cursor for 30 seconds, after which the subject heard music through their head phones. They were instructed to stop a few seconds after the music began (Wegner and Wheatley 1999, p. 488). During the movement subjects heard names of items that either appeared or did not appear on the screen. The subject was allowed to make the stop themselves on all but four trials. On these trials, the confederate would force a stop on one of the items on the screen (e.g. the swan). These stops were timed so as to be 30 s before, 5 s before, 1 s before, or 1 s after the subject heard the name of the item being stopped on (e.g. the swan; Wegner and Wheatley 1999, p. 488).

After each stop, subjects rated on a 100-point scale how much they intended to make the stop, they were told that 0 meant they merely allowed the stop to happen whereas 100 meant they intended the stop to happen (Wegner and Wheatley 1999, p. 488). Subjects gave the same scores (on average around 56) for the extent to which they intended to make the stop for when the confederate allowed them to make the stop and when the confederate forced the stop but when the stop location was named 5 s or 1 s before the stop was made. This is significantly more than the ratings given for forced stops when the name of the stop point was given 30 s before or 1 s after the stop (Wegner and Wheatley 1999, p. 489)². This suggests that, in this condition, subjects may feel they contribute to an action merely due to the presence of a prime.

There is a potential problem interpreting these results. Subjects in this study are asked to rate the extent to which they intended to make the stop. This may or may not track the subject's sense of agency. The worry being that it is asking for a judgement about the subject's mental states, rather than the action of the self. As

¹ See e.g. (Aziz-Zadeh et al. 2006) for the plausibility of such a view.

² One potential explanation for this result is that subjects attempted to stop on the named item during the forced stop conditions. If this is the case, then the higher scores would reflect the attempt to do this. Wegner and Wheatley investigated this possibility, but concluded that, while subjects reported performing visual searches for the named items, they did not seem to be making any attempt to stop on them during the forced stop trials as they made no attempt to stop on them during the free stop trials (Wegner and Wheatley 1999, p. 489).

such, on Wegner and colleagues model, it is a judgement about the mental state which is inferred to cause the action rather than about the sense of agency which is supposed to be the outcome of such an inference. This data may thus speak to judgements of apparent mental state causation and not the sense of agency at all. Perhaps, more problematically, on Wegner and colleagues' model, asking directly about the subject's intentions could alter their sense of agency, by increasing attention of the subject to the intention and thus making it more likely to be inferred to be the cause of the action. I take it here that Wegner and Wheatley assume that judgements of intention track judgements of one's own causal involvement in producing an outcome. Pretheoretically, this would also seem to track the sense of agency; however, it is not obvious on Wegner and colleagues' model that this would be true, and it remains to be verified whether judgements about intentions and judgements about agency track the same experiences.

However, if it were to be shown that this setup does track the sense of agency (i.e. that for naïve subjects, questions about intentions and agency are equivalent), then, the inference to apparent mental state causation can explain this pattern of results. The first fact to be explained is that subjects give relatively low scores of their sense of acting intentionally when they are the sole agent of action. This is explained by the fact that the exclusivity principle is not met. They believe the confederate is also deciding where to stop; thus, their intention to stop at a particular location is not the exclusive cause of the outcome. Next, we must explain why priming the finishing location causes subjects to feel they are as involved in stopping even though the stop is forced by the confederate. Here, the inference to apparent mental state causation suggests that the subject mistakenly infers that the prime is involved in causing the stop as it is prior to and consistent³ with the location of the stop. Again, as the principle of exclusivity, the subject does not feel as though they are the only agent of the action. Finally, it is a violation of the principle of priority that explains the lower ratings given when subjects heard the name of the stop location 1 s after the stop.

These findings were expanded on by Aarts et al. (2005) "wheel of fortune" task. In this study, subjects sat facing a computer screen showing eight white squares. Two grey squares moved around these squares. One square moved anticlockwise and was controlled by the subject pressing and holding the "s" key. The subjects were told that they could stop this square by pressing the "enter" key. The other grey square moved clockwise and was controlled by the computer. After some time, the subject received a cue to stop movement. At the time the stop cue appeared, the grey squares disappeared. After the subject pressed the "enter" key, one of the white squares turned black. The subject was told that this black square represented the location of either their grey square or the computer's at the point the "enter" key was pressed. In the conditions I am interested in here, subjects did not control the point where the black square appeared; instead, it always appeared four spaces ahead of where the subjects' grey square was when the stop cue appeared, i.e. four spaces

³ The notion of consistency may be problematic here. Unlike the helping hands study, these subjects were not receiving instructions to act in a particular way. Rather, they heard only the name of the object being stopped on. Such an experience is neither consistent nor inconsistent with actually stopping on that object (Bayne 2006; thanks to an anonymous reviewer for making this point.). I take it that all Wegner and Wheatley are suggesting is that the thought represents some feature of the outcome of the action. Whether or not this is a strong enough notion of consistency is a matter for further debate.

ahead of the final seen location of the subjects' grey square. After each run, subjects were asked to rate (on a 1–10 scale) the extent they felt they had controlled the stop point (Aarts et al. 2005, pp. 443, 445, 446).

In experiments 1 and 2 of this study, four conditions were compared. In experiment 1, these were (1) a subliminal prime task where the location of the black square was to appear flashed black for 0.034 s 0.046 s before the stop cue appeared, (2) the “conscious goal task” where subjects were told to try to stop their square on the square that would turn black and (3) a baseline task where no prime or instruction was given. Importantly, in the conscious goal condition (as with the other conditions), subjects were instructed to press the “enter” key as soon as possible after the stop cue appeared. If subjects adhered to this instruction, then, their “goal” to stop at a particular location should not have influenced their action. There is some reason to suppose that subjects did follow this instruction as there was no difference between reaction times for the prime and goal conditions (although both were longer than baseline), suggesting subjects did not adjust their reactions in order to hit the goal square (Aarts et al. 2005, p. 448). In experiment 2, the conscious goal condition was replaced by four, a supraliminal prime condition where the square which would turn black flashed black for 0.068 s 0.012 s prior to the stop signal appearing (Aarts et al. 2005, p. 450).

Subjects gave higher ratings for their sense of agency over the position of the black square for subliminal prime, supraliminal prime and conscious goal conditions than for baseline (Aarts et al. 2005, pp. 447, 450). There was no difference in the ratings given for either prime condition or the conscious goal condition (Aarts et al. 2005, pp. 447, 450). This result is especially interesting as it was less likely that participants stopped their square on the square that turned black in these conditions due to their increased reaction times from baseline (Aarts et al. 2005, p. 448). Thus, it seems it is possible to increase subjects' sense of agency over an action when it is, in fact, less likely that they controlled the action. These findings are explained on Wegner and colleagues' model by the subject mistakenly inferring that their “goal” or the prime caused the square to stop at a particular location. As with the helping hands and I-spy experiments, subjects do not give maximum scores on the questionnaires as the principle of exclusivity is not met. The subjects believe the final location is either a result of their action or of the computer. However, exactly what this method probes is called into doubt by a recent replication which suggests the effect of priming is specific to women (Jones et al. 2008).

Each of the studies discussed so far are taken to provide some confirmatory evidence for Wegner and colleagues' model of how the sense of agency is elicited. In particular, these studies are taken as showing that merely having a thought that is consistent with the outcome of an action prior to the occurrence of an action contributes strongly to a sense of agency over that action being elicited. That is, the principles of consistency and priority seem to hold. These studies do not speak directly to the principle of exclusivity as in all of these studies subjects have good reason to suppose that their thoughts are not the only possible cause of the action. The relevant thought can be an intention (as in the conscious goal condition above), but it can also be an auditory (I-spy and helping hands) or visual (supraliminal priming condition) experience of the outcome of the action. It may also be a subliminal perception of the outcome (subliminal priming condition). So, as well as

providing some confirmatory evidence for Wegner and colleagues' model, these studies have also been informative as to how general the relevant thought can be; it need not be an intention nor need the subject be aware of it.

The comparator model

An alternative explanation of the sense of agency comes from the comparator model advocated by Frith and colleagues (Blakemore et al. 2002; Frith et al. 2000). On this model, it is proposed that the sense of agency is elicited, not by an inference to apparent mental state causation, but by the motor control system. In particular, it is proposed that the sense of agency is elicited when the actual sensory consequences of an action match the motor control systems prediction about what will happen based on the motor commands produced. It is a represented match between the actual and predicted sensory outcomes of action that elicits the sense of agency.

A strength of Wegner and colleagues' model is the relative difficulty the comparator model has with explaining the above results⁴. It is difficult to see how the comparator model could explain the results of I-Spy, helping hands and the wheel of fortune. Take helping hands for example. Subjects do not perform the action they see; thus, it seems unlikely that they would get a positive match for an actual/predicted sensory feedback comparison (Synofzik et al. 2008, p 226). As such, it is not obvious how the comparator model could explain these data.

A potential explanation arises from the possibility that subjects may perform slight movements consistent with the heard instruction in the helping hands study. Indeed, being told to imagine a certain movement causes muscle contractions consistent with that movement (Wegner et al. 2004, p 847). Could these movements and the accompanying comparisons cause a sense of agency over the seen movement of the arms? Wegner et al. (2004) tested this possibility by examining the effect of performing specific movements on the sense of agency ratings given. Subjects were told either to mimic the movement described whilst still keeping their arms by their side or to perform a distracting movement, i.e. tapping the side of their legs (Wegner et al. 2004, pg 843-844). Those who were instructed to mimic the action described gave higher ratings for their sense of agency over the action seen than those who were given no instruction (Wegner et al. 2004, p 844). This suggests that performing a movement similar to that which is seen may increase one's sense of agency over the seen action. However, there was no difference between the sense of agency ratings given by those performing distracter actions and those given no instruction (Wegner et al. 2004, pg 845). This suggests that the seen action does not need to match the predictions being made by the motor system in order for one to experience a sense of agency over it; hence, it is not clear how the comparator model can explain this result.

The wheel of fortune and I-spy experiments have also been used to challenge the comparator model as an account of the sense of agency (Synofzik et al. 2008, p 226). The conscious goal condition of the wheel of fortune study seems possible to explain

⁴ Ultimately, I contend the comparator model has a lot going for it and may be able to explain these studies (Carruthers 2009; Carruthers 2010). The point here is that it is relatively easier for Wegner and colleagues' model to explain these results.

on the comparator model. Perhaps, the subject produces a prediction from the goal of what will happen from the goal to stop on a particular square. This prediction can then be compared with the perceived location of the black square. In cases of a match, the sense of agency would be elicited. However, the priming cases in both studies seem harder to understand. Why should priming a location influence the sense of agency? In other words, how could a subliminal or supraliminal auditory or visual prime cause the sense of agency to be elicited? There is no straightforward answer to this on the comparator model. However, as we have already seen, Wegner and colleagues' model can explain these results. As the prime is prior to the action and consistent with its outcome (in the sense that it represents the same stopping location), it is mistakenly inferred to cause the action of stopping at that location. As the principle of exclusivity is violated in these cases, the action is the subject's, a confederate's, or a computer's; subjects do not respond with certainty on the scales measuring the extent to which they experience themselves as agents. As such, it seems that the inference to mental causation offers a powerful explanation of the sense of agency and its experimental manipulations.

The sense of agency without an inference to mental state causation

In this section, I consider an empirical problem for Wegner and colleagues' model. A prediction of this model which has received little attention is that if the sense of agency is elicited by an inference to mental state causation, then, this sense should be tied closely to performance on certain theory of mind tasks, in particular, those that measure one's understanding of the relationship between one's own intentions and actions. The data I consider suggests that children can experience a sense of agency even when they are unable to make inferences to their mental states as causes of their actions. I consider a variety of possible cases that can be posed from the literature. Many of the examples face problems of their own. The example from Nahmias (2005) of playing sport in the zone which I discuss below has yet to be systematically investigated, and many of the other studies, involving healthy children and those suffering from autism, seem to inappropriately depend on children's memory rather than experience of their own intentions. Despite this, there is strong evidence that children experience a sense of agency before being able to infer that their mental states cause their actions. Furthermore, there is evidence that children suffering from autism experience a sense of agency without being able to infer that their mental states cause their action.

Sense of agency without an inference to mental state causation in healthy adults and normally developing children

One argument against the claim that experiences of agency are dependent on inferences about mental state causation is due to Nahmias. He points out that it is a prediction of Wegner and colleagues' description that actions that are not experienced as occurring after a thought about that action will not be experienced as due to oneself (Nahmias 2005, p. 783). This is a prediction that Wegner and colleagues embrace (Wegner and Wheatley 1999, p. 848). A well-known example of

an action that appears not to be preceded by thoughts about the action is playing sport “in the zone” (Nahmias 2005, p. 783). Nahmias takes it that such actions are experienced as arising from oneself despite the fact that they are not experienced as arising from any particular mental state, i.e. the athlete does not know their intention-in-action. This suggests dissociation between the sense of agency and the possibility of making the inference that an intention causes one’s action. If he is right, then professional athletes should report that they perform their actions on the field, but that they do not experience an intention to perform any particular action. However, if Wegner and colleagues are right, then they should report that they do not perform their actions on the field. I know of no study that addresses this directly, nor is it obvious how this could be assessed independently of athlete’s memories of particular events. However, there are studies which suggest that the sense of agency can be elicited even when it is not possible for one to infer that one’s intention caused one’s action.

Montgomery and Lightner (2004) administered a detailed and elegant series of tests to 3- and 4-year-olds. Here, I discuss just two of their results, those that demonstrate the problem for Wegner and colleagues’ model in the most straightforward way. In one study, children were shown a picture, say of a ball, which was to be copied in a variety of ways. First, the child sat with their eyes closed while the experimenter moved the child’s hand in order to create the picture. Second, after the drawing was made in this way, the child watched as the experimenter changed the drawing to a different object (e.g. changed a picture of a ball to a picture of a clock). In the third condition, the child produced a picture themselves, e.g. they copied the picture of a ball. In the fourth condition, the child produced a copy themselves and then watched the experimenter alter the drawing. After each drawing was made, the children were asked who made the final drawing. The vast majority of 3- and 4-year-olds easily identified who was responsible for the final drawing in each case (Montgomery and Lightner 2004, pp. 428–429). That is, 3- and 4-year-olds can identify the agent behind the drawing. This suggests that they have a sense of agency. However, children of the same age are much worse at identifying relationships between their thoughts and their actions (in this case their drawings). Despite the fact that they were correctly able to identify the experimenter as the agent behind the altered drawings, they nevertheless claimed that they had tried to produce the final (altered) picture. In conditions where they drew one object (a ball) that was then altered by the experimenter (to be a clock), children claimed that they had tried to draw the latter object (a clock) even though they could identify the experimenter as the agent behind the latter drawing (Montgomery and Lightner 2004, pp. 425–427).

Based on this study, it seems that 3- and 4-year-olds are capable of identifying the agent of an action. This suggests that they have a sense of agency to use in making these judgements. However, the same children are poor at recognising the relationship between their intentions and their actions. They believe that they intended the final outcome of an action even when they could not have done so. As such, they do not have the resources to make an inference to their intentions as the cause of their actions. Such children do not know their intentions, and thus, they cannot infer that their intention caused their action. This suggests that the sense of agency can be elicited independently of the child’s ability to make an inference to

apparent mental state causation. As such, this inference cannot always be the basis for the sense of agency.⁵

Sense of agency without an inference to mental state causation in children suffering from autism

Wegner and colleagues' model makes a further prediction that is problematic. Wegner (2002) claims that "there may even be people who regularly fail to grasp the consistency between their thoughts and their actions and thus experience the actions as less than willed" (Wegner 2002, p. 133). Children suffering from autism would seem to fit the criterion of having a deficit in the ability to recognise the relationship between their intentions and their actions. This is recognised by Wegner and Wheatley when they suggest that the sense of agency comes from taking the intentional stance toward oneself (Wegner and Wheatley 1999, p. 490). One of the key deficits in autism is difficulty in taking the intentional stance toward others and oneself (see e.g. Frith and Happe 1999; Baron-Cohen et al. 1985; Baron-Cohen 1995; Frith 2003; Gopnik 1993). However, it is probable that people suffering from autism have an intact sense of agency.

Children suffering from autism are worse at identifying the relationship between their intentions and actions than healthy controls. This is suggested by a comparison of performance on a target shooting game. The aim of the game was to select one target of four and try to hit it. One of the four targets contained a prize. When the outcome of the shot is rigged so that the children always hit the target that contains the reward (regardless of what they aimed at), healthy 5-year-olds claimed that they had intended to hit another target even though their intention to win the prize had been satisfied. This was much harder for those with autism who consistently claimed to have intentionally hit the target containing the prize (Frith and Happe 1999, p. 9; Russell and Jarrold 1998, p. 179). This suggests that those with autism have a deficit in understanding the relationship between their intentions and the outcome of actions. As with the young children in Montgomery and Lightner's study these patients believe that they intended the outcome that actually occurred even when this is not the case. As such, on Wegner and colleagues' model, these children should also experience a deficit in their sense of agency.

Williams and Happe (2009a) investigated the ability of children suffering from autism to recognise their own intentions. They did so by examining if such children

⁵ At this point, the advocate of Wegner's model may be tempted to propose that studies like those from Montgomery and Lightner simply probe a different phenomenon to what the inference to mental causation is supposed to explain. It may be the case that they measure judgements of agency and not the sense of agency. Might subjects in this study be forming judgements that they are agents of the action without having a feeling of doing so? This seems unlikely. It is not even clear what it would mean to form a judgement of one's own agency without having any sense of agency. Although judgements of one's own agency are not identical to the sense of one's own agency, it seems that they are usually the sense of agency plus. That is, they are the sense of agency plus some conceptual level judgement. The advocate of this objection would to show that judgements of one's own agency can occur without having a sense of agency and that this is what is happening in these cases. The same applies to the cases discussed below where the sense of agency is operationalised via some judgement. It also applies to studies such as "helping hands" and "I-spy" which operationalise the sense of agency via judgements of agency and were used to support Wegner and colleagues model. However, this does not apply to cases where the sense of agency is operationalised via something other than judgements of agency, in particular, those where the sense of agency is measured via self-monitoring performance in e.g. the squares task.

could state whether or not a movement or action had been intentional in two tasks. The first task, called the knee-jerk task, examined subjects ability to recognise a knee-jerk reflex as unintentional. Subjects had their knee struck with a hammer usually used to examine this reflex. When the reflex was elicited, the experimenter said: "Look, your leg moved. Did you mean to move your leg?" (Williams and Happe 2009b, p. 5). The correct answer is, of course, no. Those suffering from autism found this task harder than those in a control group (either learning disabled or healthy children). Only 68.2% of those suffering autism passed this test; in contrast, 95.5% of those in the control group passed (Williams and Happe 2009b, p. 5).

Williams and Happe (2009b) expanded on these findings using the "transparent intentions" task. This task is reminiscent of the task from Montgomery and Lightner (2004) discussed above. Subjects were given two transparencies of the type used with overhead projectors. Each transparency had an incomplete picture on it, say a choir boy missing an ear or a cup missing a handle. When placed on top of one another, only the image on the bottom sheet would be visible (e.g. the boy missing an ear). In one case, the incomplete picture of the coffee cup is placed on top of the picture of a choir boy such that only the choir boy can be seen. In the second example, the unfinished picture of an ice cream is placed on top of the unfinished picture of a tree such that only the tree is visible.

Subjects were presented the transparencies together in this form and asked to name the missing part of the image, e.g. the boy's ear or some of the leaves on the tree, and then draw it. After the drawing was made, the transparencies were separated so the subject could see both images. The drawing they added now completed the image they had not previously seen. That is the curve made to represent the boy's ear now appeared as the handle on the cup and the series of curves made to represent leaves now appeared to be ice cream on top of the cone.

Subjects were then asked three questions. The control question was what they had drawn (the handle on the cup or ice cream). The two test questions probed what the subjects meant (intended) to draw (an ear on the boy or leaves) and what they thought they were drawing (again an ear on the boy or leaves; Williams and Happe 2009b, p. 8). Those children suffering autism were as good as controls on the control and think questions, but significantly worse of the mean question (Williams and Happe 2009b, p. 9). These two tests seem to suggest that children suffering from autism are worse at recognising their own intentions than healthy children or those suffering from learning disabilities. However, most children with autism (around 70%) do pass these tests. So, it seems that there is a subset of children suffering autism who lack access to their own intentions when forming judgements about what had caused their actions. If Wegner and colleagues' model is right, these patients should have a deficit in their sense of agency, that is around 30% of children suffering from autism should have a deficit in their sense of agency. As a group, children suffering from autism should do worse on measures of the sense of agency than healthy controls or those suffering from learning disabilities.

Although they did not test for an association between deficits in understanding intention and the sense of agency, there are studies that suggest that children with autism show the same performance as controls on tasks that measure their sense of agency. This seems to be suggested by a study on which children with autism are able to identify the agent behind an action to the same extent as controls. In a study by Hill and Russell (2002), children were presented with pairs of familiar objects

(say a coin and a candle) with which either they or an experimenter performed an action (put the coin under the candle). After 18 trials, subjects were given a memory test which included all 18 pairs and nine new combinations. Subjects were asked who has performed the actions they correctly remembered. It was found that those with autism were as good as those with a moderate learning disability and healthy controls in identifying the agent behind the action (Hill and Russell 2002, p. 164).

Farrant et al. (1998) performed a similar test. Experimenter and subject sat together and listened to a recording that asked them to say a set of words that should be familiar to young children. After this, the children were given a memory test which involved presenting them with a new list of words. Children were asked who had said the words they correctly identified as having being said aloud in the first phase. Children with autism were as good at answering this question as mental aged-matched children with mild mental retardation and healthy controls (Farrant et al. 1998, pp. 47–48). This suggests that children with autism do not suffer a deficit in their sense of agency. However, this finding is not ideal as the subject and the experimenter each held different-coloured blocks as identifying marks. As the authors note, this creates the possibility of children learning to associate particular words with a particular coloured block and not using their sense of agency to judge who said what (Farrant et al. 1998, p. 48).

Whilst suggestive, these studies are not perfect for my purposes as they involve memory of the agent of an action (see Russell and Jarrold 1999; Williams and Happe 2009a experiment 2 for other measures of the sense of being an agent across time). There may also be a problem relating to using judgements of agency to operationalise the sense of agency. To speak directly to the sense of agency, it is better to use tasks along the lines of the action-monitoring and self-recognition tasks used to measure the sense of agency as it occurs. Performance on these tasks suggests that those suffering autism have a normal sense of agency over their bodily actions.

One reason to suppose that subjects suffering from autism have an intact sense of agency is that they pass self-recognition tests that seem to require such a sense. Recently, Williams and Happe (2009a) have examined the self-recognition/action-monitoring abilities of those suffering from autism. In this study, Williams and Happe compared the performance of children suffering from autism to those with learning disabilities on an action-monitoring task called “the squares task”. For this task, subjects sat facing a computer screen which displayed a variable number of squares. When the mouse was moved, all of the squares would move. One square, called the target square, moved consistently with the movements of the mouse. All the other squares, called distracter squares, moved with various degrees of variation from the mouse (Williams and Happe 2009a, pp. 253–254). The subjects' task was to identify the target square. The mouse was obscured inside a box so that subjects could not visually match the movement of the mouse to that of the target square. The difficulty of this task was altered by varying the number of squares on the screen between four and 25 and by varying the possible paths the distracter squares could take to 360°, 180°, and 90° from the trajectory of the target (Williams and Happe 2009a, p. 254). Subjects were tested under two conditions. These are (1) the self-task where subjects themselves moved the mouse and (2) the other task where subjects passively held onto the mouse while the experimenter moved it. Those with a deficit in their sense of agency should find the self-task harder than controls, and they

should show no difference in their performance on the self and other tasks. In contrast, those with a sense of agency should find the self-task easier than the other task as they have their sense of agency to aid in making the match.

The results of the study suggest that those suffering from autism have a normal sense of agency. Subjects suffering from autism as well as controls both found the self-condition easier than the other condition. Similarly, there was no difference between the performance of those suffering from autism and controls on the self-condition (Williams and Happe 2009a, p. 255). This suggests that those suffering from autism have the same sense of agency as normally developing children. This falsifies the prediction of Wegner and colleagues' model.

These findings suggest that children experience a sense of agency even when they lack the ability to identify a causal relationship between their own intentions or "tryings" and action. They experience a sense of agency even when they are not capable of making an inference to apparent mental state causation. Young children and those with autism who are unaware of their intentions can still identify the agent of an action with ease and perform well on action-monitoring tasks that require a sense of agency. This is problematic for the hypothesis that inferences to mental state causation underlie the sense of agency. However, it cannot form an objection to that model until a viable alternative explanation is presented. We could pursue this in several ways; say, by showing how the comparator model might explain Wegner and colleagues' results (see e.g. Carruthers 2010) or by developing a new model (see e.g. Synofzik et al. 2008). This is a task for another time.

Conclusion

In this paper, I have argued that Wegner and colleagues model of how the sense of agency is elicited faces a problem in that it appears possible to experience a sense of agency even when one cannot make inferences to apparent mental state causation. That is, it is possible to experience a sense of agency behind an action without being able to recognise the relationship between one's intentions and one's actions. This is inconsistent with Wegner and colleagues' model which attempts to explain the sense of agency in terms of inferring one's intentions are the cause of one's actions. Instances of this phenomenon can be seen in normally developing children and those suffering from autism.

Acknowledgements This work was undertaken toward a Ph.D. at Adelaide University and Macquarie University. I thank John Sutton, Philip Gerrans, Max Coltheart and Gerard O'Brien for their work in supervising my thesis. My thanks go to two anonymous reviewers for their detailed critiques. Thanks also go to the audiences at the Berlin School of Mind and Brain and the Association for the Scientific Study of Consciousness for their stimulating discussions of previous versions of this paper. Finally my thanks go to Elizabeth Schier and Kristina Musholt for their extensive comments on early drafts.

References

- Aarts, H., Custers, R., & Wegner, D. M. (2005). On the inference of personal authorship: enhancing experienced agency by priming effect information. *Conscious Cogn*, 14, 439–458.
- Aziz-Zadeh, L., Wilson, S. M., Rizzolatti, G., & Lacoboni, M. (2006). Congruent embodied representations for visually presented actions and linguistic phrases describing actions. *Curr Biol*, 16, 1818–1823.

- Baron-Cohen, S. (1995). *Mindblindness: an essay on autism and theory of mind* (p. 168). Cambridge, MA: MIT Press.
- Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind"? *Cognition*, *21*, 37–46.
- Bayne, T. (2006). Phenomenology and the feeling of doing: Wegner on the conscious will. Does consciousness cause behaviour? An investigation of the nature of volition. In S. Pockett, W. P. Banks, & S. Gallagher (Eds.), *MIT Press*. MA: Cambridge.
- Blakemore, S. J., Wolpert, D. M., & Frith, C. D. (2002). Abnormalities in the awareness of action. *Trends Cogn Sci*, *6*(6), 237–242.
- Carruthers, G. (2009). Commentary on Synofzik, Vosgerau and Newen 2008. *Conscious Cogn*, *18*(2), 515–520.
- Carruthers, G. (2010). A comparison of fortunes: the comparator and multifactorial weighting models of the sense of agency. *9th Meeting of the Australasian Society for Cognitive Science*, Macquarie University, Sydney
- Farrant, A., Blades, M., & Boucher, J. (1998). Source monitoring by children with autism. *J Autism Dev Disord*, *28*(1), 43–50.
- Frith, C. D., Blakemore, S.-J., & Wolpert, D. M. (2000). Abnormalities in the awareness and control of action. *Philos Trans R Soc Lond B Biol Sci*, *355*, 1771–1788.
- Frith, U. (2003). *Autism: explaining the Enigma* second edition, Blackwell.
- Frith, U., & Happe, F. (1999). Theory of mind and self-consciousness: what is it like to be autistic? *Mind Lang*, *14*(1), 1–22.
- Gopnik, A. (1993). How we know our minds: the illusion of first-person knowledge of intentionality. *Behav Brain Sci*, *16*, 1–14.
- Hill, E., & Russell, J. (2002). Action memory and self monitoring in children with autism: self versus other. *Infant Child Dev*, *11*, 159–170.
- Jones, S. R., de-Wit, L., Femyhough, C., & Meins, E. (2008). A new spin on the wheel of fortune: priming of action-authorship judgements and relation to psychosis-like experiences. *Conscious Cogn*, *17*, 576–586.
- Montgomery, D. E., & Lightner, M. (2004). Children's developing understanding of differences between their own intentional action and passive movement. *Br J Dev Psychol*, *22*, 417–438.
- Nahmias, E. (2005). Agency, authorship and illusion. *Conscious Cogn*, *14*, 771–785.
- Russell, J., & Jarrold, C. (1998). Error correction problems in autism: evidence for a monitoring impairment? *J Autism Dev Disord*, *28*(3), 177–188.
- Russell, J., & Jarrold, C. (1999). Memory for actions in children with autism: self versus other. *Cogn Neuropsychiatry*, *4*, 303–331.
- Spanos, N. P. (1982). Hypnotic behaviour: a cognitive, social psychological perspective. *Res Commun Psychol Psych Behav*, *7*, 199–213.
- Stephens, G. L., & Graham, G. (2000). *When self-consciousness breaks: alien voices and inserted thoughts*. Cambridge, MA: MIT.
- Synofzik, M., Vosgerau, G., & Newen, A. (2008). Beyond the comparator model: a multifactorial two-step account of agency. *Conscious Cogn*, *17*, 219–239.
- Wegner, D. M. (2002). *The illusion of conscious will*. Cambridge, MA: MIT Press.
- Wegner, D. M. (2003). The mind's best trick: how we experience conscious will. *Trends Cogn Sci*, *7*(2), 65–69.
- Wegner, D. M., Sparrow, B., & Winerman, L. (2004). Vicarious agency: experiencing control over the movements of others. *J Pers Soc Psychol*, *86*(6), 838–848.
- Wegner, D. M., & Wheatley, T. (1999). Apparent mental causation: sources of the experience of will. *Am Psychol*, *54*(7), 480–492.
- Williams, D., & Happe, F. (2009a). Pre-conceptual aspects of self-awareness in autism spectrum disorder: the case of action monitoring. *J Autism Dev Disord*, *39*, 251–259.
- Williams, D., & Happe, F. (2009b). Representing intentions in self and other: studies of autism and typical development. *Dev Sci*, 1–13.