Neuroscience of Consciousness, 2019, 5(1): niz006

doi: 10.1093/nc/niz006
Opinion Paper

A match does not make a sense: on the sufficiency of the comparator model for explaining the sense of agency

Lorijn Zaadnoordijk[1,2,*,†], Tarek R. Besold[3,‡] and Sabine Hunnius[1]

[1]Donders Institute for Brain, Cognition, and Behaviour, Radboud University, Nijmegen, The Netherlands; [2]Trinity College Institute of Neuroscience, Trinity College Dublin, Dublin, Ireland; [3]Alpha Health AI Lab, Telefonica Innovation Alpha, Barcelona, Spain

*Correspondence address: Trinity College Institute of Neuroscience, Trinity College Dublin, Dublin, Ireland. E-mail: L.Zaadnoordijk@tcd.ie

†Lorijn Zaadnoordijk, https://orcid.org/0000-0002-7484-6995

‡Tarek Richard Besold, https://orcid.org/0000-0002-8002-0049

Abstract

The development of a sense of agency is indispensable for a cognitive entity (biological or artificial) to become a cognitive agent. In developmental psychology, researchers have taken inspiration from adult cognitive psychology and neuroscience literature and use the comparator model to assess the presence of a sense of agency in early infancy. Similarly, robotics researchers have taken components of the proposed mechanism in attempts to build a sense of agency into artificial systems. In this article, we identify an invalidating theoretical flaw in the reasoning underlying this conversion from adult studies to developmental science and cognitive systems research, rooted in an oversight in the conceptualization of the comparator model as currently used in experimental practice. In these experiments, the emphasis has been put solely on testing for a match between predicted and observed sensory consequences. We argue that the match by itself can exclusively generate a simple categorization or a representation of equality between predicted and observed sensory consequences, both of which are insufficient to generate the causal representations required for a sense of agency. Consequently, the comparator model, as it has been described in the context of the sense of agency and as it is commonly used in experimental designs, is insufficient to generate the sense of agency: infants and robots require more than developing the ability to match predicted and observed sensory consequences for a sense of agency. We conclude with outlining possible solutions and future directions for researchers in developmental science and artificial intelligence.

Key words: sense of agency; infancy; developmental cognitive science; developmental robotics; comparator model; artificial intelligence

Introduction

It is common practice in fields such as developmental science and robotics to use models proposed in cognitive (neuro)science as a starting point for setting up experiments to test for specific cognitive capacities or implement them artificially (e.g. Banks and Salapatek 1981; Johnson 1997; Burghart *et al.* 2005; Shanahan 2006; Qiao *et al.* 2016). However, an extrapolation of theories and measures designed for non-clinical adults to infants or robots has important caveats. One weakness is that the transfer presupposes that the mechanisms captured by the

Received: 16 October 2018; **Revised:** 23 March 2019. **Accepted:** 30 March 2019

models used in adult research are "explanatory" of the capacity in general—i.e., that a model can, through its components and the interaction of processes between these components, account for the entire phenomenon of interest (Bechtel 2009)—rather than "descriptive" of a condition under which the capacity might occur. In this article, we demonstrate the consequences of this particular caveat using the sense of agency (i.e. the experience of oneself as a causal agent) as example for a cognitive capacity. Our analysis focuses on the mechanism proposed in the comparator model as a suggested mechanism within the context of a sense of agency in particular.

The type of analysis we are undertaking is relevant for any population in which the presence of the full range of cognitive capacities of an adult human cannot be automatically assumed, either because they might not yet have developed or matured (as is the case, for instance, in infants), or because they have not yet been successfully implemented (as might be the case in artificial systems). For both types of populations, similar caution is required with regards to whether mechanisms originally proposed on the basis of adult research can account for the entire phenomenon. A cognitive process in adults may draw upon additionally available capacities, whose presence cannot tacitly be assumed in the aforementioned populations—introducing a need to explicitly account for them in the corresponding models. In that light, in what follows, the comparator model—or, more precisely, the way in which it is used in experimental and implementational practice—is put under scrutiny. We demonstrate that without sufficient evaluation as to whether the mechanism in its proposed form can entirely explain, and thus generate, a sense of agency, interpreting the evidence for this mechanism in infants in the same way as in an adult population can lead to incorrect conclusions. Moreover, the explanatory value of the mechanism directly reflects upon the usefulness of implementing the mechanism in an artificial agent in order to bring about the phenomenon of interest. Assessing the explanatory value of a mechanism is thus essential to ensure valid conclusions and to provide solid foundations for subsequent research lines.

The Comparator Model and Its Application in Research

The sense of agency refers to the experience of oneself as an agent who can cause events by acting. This experience is closely linked to the ability to distinguish events caused by one's own actions from those caused by other agents or external forces (see for review David *et al.* 2008; Chambon *et al.* 2014; see for meta-analysis of the neural underpinnings Sperduti *et al.* 2011). The experience of agency in turn is an essential aspect of social behavior, self-awareness, and causal learning (Lagnado and Sloman 2002; David *et al.* 2008; David 2012). It has been postulated that the ability to attribute events to oneself or to others can be explained by the comparator model, which is rooted theories of motor control (Blakemore *et al.* 2002; see also Fig. 1). For action selection and action awareness, two types of internal models are used: an inverse model to select the action that will (most likely) lead to the desired goal and a forward model to monitor the ongoing process and its final result by comparing the sensory information to the predicted state (see Wolpert and Kawato 1998 for a formalized description of these processes). The sensory prediction is thought to be based on the efferent signals (i.e. the motor command) and is compared to the afferent sensory signals. According to the comparator model theory of the sense of agency, whenever the prediction and the actual outcome "match" (i.e. are congruent), it is assumed that this sends a cue for people to experience a sense of agency (e.g. Gallagher 2000; Jeannerod 2009). The experienced agency is largely dependent on the degree of congruence versus incongruence between the predicted and actual sensory outcomes (Sato and Yasuda 2005).

Before assessing the comparator model as an explanatory model for the sense of agency, it is useful to understand its background. The comparator model was not originally suggested as a model for sense of agency but as a physiological mechanism of sensorimotor control. As early as 1950, two studies were published demonstrating the role of efference signals (also known as efference copy or corollary discharge) in monitoring and optimizing motor control (Sperry 1950; Von Holst and Mittelstaedt 1950). A comparison between efference (i.e. motor) and reafference (i.e. sensory) signals was proposed as a physiological cue to classify signals as self-produced or externally produced. These empirical findings have been further investigated and theoretically expanded (e.g. Wolpert and Kawato 1998). Currently, a large body of empirical evidence supports the comparator model as a model for mechanisms of sensorimotor control (e.g. Miall and Wolpert 1996; Sabes 2000; Wolpert and Flanagan 2001; Shadmehr *et al.* 2010).

More recently, the comparator model has been extended to also serve as a neurocognitive model of the sense of agency (Frith *et al.* 2000; Gallagher 2000; Frith 2005; Jeannerod 2009). The validity of this claim has been tested in many psychological and neuroscientific experiments with adults (e.g. Fourneret and Jeannerod 1998; Farrer *et al.* 2003, 2004). In these studies, participants perform an action while monitoring the effects on a computer screen. The observed effect is manipulated such that it violates the temporal or spatial contingency to the participant's action, and participants are asked to judge whether they produced the observed effect. They tend to judge the effects as being externally generated (i.e. they feel reduced or no agency regarding the movement) more often when the manipulation is stronger. This judgment is commonly thought to be made through assessing the action-outcome contingency by comparing the predicted sensory consequence based on the efferent signals to the observed sensory feedback based on the reafferent signals. Moreover, testing for the importance of efferent signals, Tsakiris *et al.* (2005) found that when the participants' fingers were being moved and only visual afferent signals were available, participants had more difficulty recognizing the displayed hand as their own, and as such their self-recognition performance decreased drastically. This was taken to be indicative of the key contribution of efferent signals (for an overview of experiments investigating the sense of agency in adults, see David *et al.* 2008).

Following adult research, the ability to detect a match between the predicted and actual sensory consequences of an action has also been taken as an indication for an emerging sense of agency in early infancy. In order to investigate infants' capacity to detect whether the consequences of an action are as they predicted, several researchers (e.g. Bahrick and Watson 1985; Rochat and Morgan 1995, 1998) conducted experiments in which infants saw their own legs projected on a screen in real time and on another screen, positioned directly next to the first one, a distorted (e.g. mirrored) projection of their legs or the projection of another infant's legs. In these studies, researchers found differences in infants' looking behavior to the contingent and non-contingent displays of their legs. These studies showed that infants as young as 3 months of age differentiate between the contingent and non-contingent displays and thus display

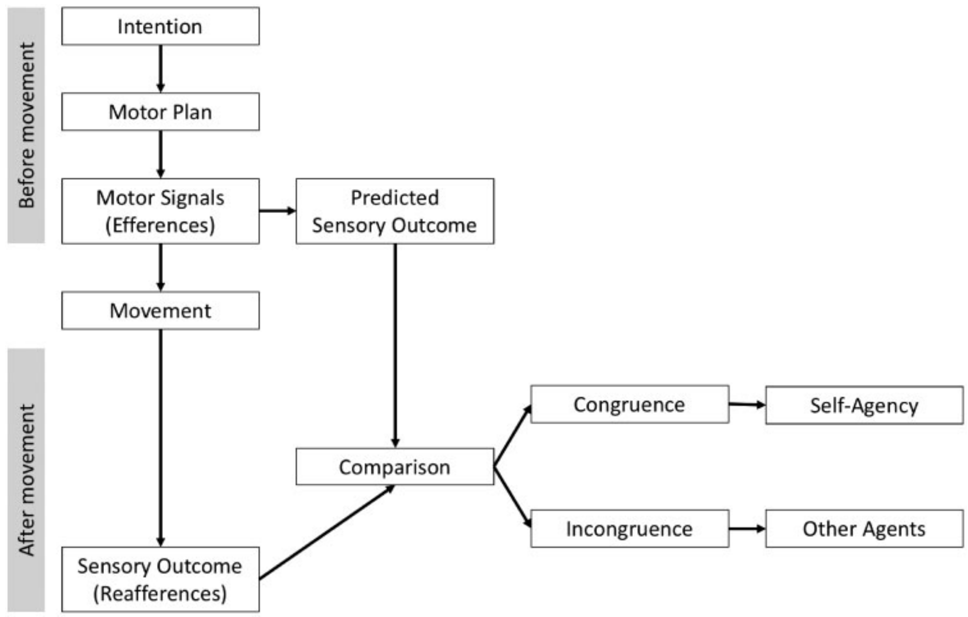

Figure 1. A schematic overview of the comparator model [adapted from David *et al.* (2008)].

sensitivity to the degree to which the sensory consequences match the motor signals they send out.

More recently, Watanabe and Taga (2011) and Kelso (2016) have argued that the developing sense of agency can be demonstrated using the so-called mobile-paradigm. In this paradigm, originally used to study memory retention, the infant's limb is tethered to an overhead mobile such that when the infant moves the connected limb causes contingent movement of the mobile (Rovee and Rovee 1969). Infants react to this contingency by increasing their movement frequency (e.g. Watson 1972; Rovee-Collier *et al.* 1978; Heathcock *et al.* 2004, 2005; Watanabe and Taga 2006, 2011), which has been taken as an indication that they have learned the causal action-effect relation (Watson 1972, 1981; Gergely and Watson 1999; Watanabe and Taga 2006, 2011; Kelso 2016). This finding of an increase in action frequency when the action elicits a contingent effect is in line with previous work by Rochat and Striano (1999), who argue that the infant's explorative reaction to contingent effects can be taken as evidence for the emergence of a sense of agency. These experiments and their interpretations serve as examples giving evidence that contingency detection has hitherto been taken as primary cognitive indicator for the sense of agency in research investigating its early emergence.

Inspired by the comparator model as a proposed mechanism for the sense of agency in developmental psychology and cognitive neuroscience, developmental robotics researchers have taken implementations of the model as a way to imbue artificial agents with a sense of agency. In an approach collapsing sense of agency, sense of body-ownership, and sense of selfhood, Pitti *et al.* (2009) equipped a head-neck-eyes robot with the ability to detect contingencies in sensorimotor networks using an artificial neural network that models spike timing-dependent synaptic plasticity (STDP) as observed in the central nervous system. STDP models the process of Hebbian learning and the constituent change in connection strength between pre- and postsynaptic neurons, taking into account the need for the presynaptic neuron to fire before the postsynaptic neuron to establish proper temporal dynamics corresponding to the ascribed causal connection. The resulting neural architecture implemented by Pitti *et al.* (2009) represented the system's self-produced visuomotor information, making the detection of sensorimotor contingencies possible by inspecting the clusters of neurons whose connections had been strengthened by a reinforcement learning algorithm. Over time, congruent sensorimotor neural pairs are reinforced, whilst incongruent ones are weakened and eventually inhibited. This gives rise to representations of sensorimotor contingencies that allow the system to anticipate ongoing sensorimotor activity and predict the system's next sensory input [albeit limiting them to what the authors call "the here and now" (Pitti *et al.* 2009, p. 87)]. As a consequence, the robot started to act upon the sensorimotor contingencies, which the authors took to represent one of the most basic levels of self-awareness and thereby, in their reading, a sense of agency (with the neural dynamics coherence, i.e., the rate of correct and comprehensive predictions of the actual state, as associated quantitative measure).

Other implementations of the sense of agency in artificial systems were based on previous findings that organisms attenuate predicted incoming sensory signals. The rationale behind sensory attenuation is that predictable signals require less attention than unexpected signals and are therefore processed differently, resulting in reduced perceptual intensity (Blakemore *et al.* 1999). Since self-produced signals tend to be maximally predictable, Schillaci *et al.* created an artificial system that could use the prediction of self-produced auditory signals in a sensory attenuation process (Bechtle *et al.* 2016; Pico *et al.* 2016; Schillaci *et al.* 2016). In their series of studies, the researchers implemented a mechanism to classify self-produced and externally produced signals in a robot. They showed that the system processes the externally produced auditory signals as more salient and uses the predictions generated by the comparator model to filter out the self-produced signals.

In sum, researchers from various disciplines have so far followed adult research in focusing on the detection of the match between the predicted and observed sensory consequences of an action as indicator for a sense of agency. However, this focus crucially relies on the tacit assumption that the sole presence of match detection is sufficient to generate the sense of agency. In the remainder of this article, we take a step back and evaluate whether the detection of a match has sufficient explanatory value to conclusively investigate the emergence of a sense of agency in infant development or as a blueprint for a sense of agency mechanism in a robot. In our analysis, we do not focus specifically on the assumptions made by the comparator model (e.g. in motor control research) but rather on how the comparator model has been applied cross-disciplinarily in experiments and robotic implementations in the context of the sense of agency. The main question for our conceptual analysis is whether the matching mechanism of the comparator model is sufficient to produce a sense of agency. As stated initially, this sufficiency is crucial, as developmental researchers and roboticists cannot take for granted at the time of the experiment that any other potentially necessary components have already developed or been built in, so as to complement the mechanism such that it carries out all requisite processes for a sense of agency.

Analysis of the Explanatory Value of the Comparator Model

It may seem intuitive to consider a mechanism that classifies self-produced from externally produced signals to be able to generate a sense of agency. However, we argue that even if a system can differentiate between signals that have been self or externally produced, this is not necessarily sufficient for a sense of agency. Moreover, we subsequently argue that even in cases where the match represents information based on the comparison between the predicted and observed consequence going beyond a simple categorization, this is still not sufficient for the model to be explanatory of the phenomenon.

Parts of this confusion may have their roots in the way the comparator model's relation to a sense of agency is verbalized. For instance, David *et al.* (2008) described the relation as follows: "Thereby, the sense of agency particularly hinges on the forward model, which uses an efference copy, that is, a copy of a motor command predicting respective sensory consequences. Accordingly, congruence of the predicted with the actual consequence, then, supposedly would lead to the attribution of the sense of agency to oneself, whereas incongruence would indicate another agent as the cause of an action." (p. 524) Although this description might be accurate (i.e. in the condition in which a match is detected, a sense of agency is assumed to occur in a population that has already developed the capacity for the experience), it may not be a description of an explanatory mechanism that contains all components required to bring about the sense of agency; it remains agnostic about the processes that give rise to it. In adults, many more processes may be involved to give rise to the experience but are implicitly assumed by the researcher as already present and, thus, are not studied. In infants however, we cannot assume all adult cognitive capacities to already have developed and thus should not be looking for a single necessary mechanism if we aim to study the presence or emergence of a capacity. Finding evidence for a necessary mechanism means obtaining evidence for an indispensable subcomponent of the capacity, but it does not automatically mean that all required processes have developed. Although the capacity of interest does not exist without the necessary mechanism, the presence of the necessary mechanism is not sufficient for the presence of the phenomenon of interest. Rather, the presence of a capacity can only be inferred from evidence for a sufficient mechanism. The comparator model would represent a "necessary" mechanism if (1) the comparator model described a condition under which the sense of agency occurs and (2) this condition must always be met for this population to experience their agency [though see Synofzik *et al.* (2008) for criticism targeting the necessity of the comparator mechanism for the sense of agency]. However, for a mechanism to be explanatory it must be "sufficient" to account for the entire phenomenon (Bechtel and Abrahamsen 2005; Bechtel 2009). That is, the mechanism must be able to produce the cognitive phenomenon of interest, in this case the sense of agency, by virtue of its processes and subcomponents. In contrast to previous work, which evaluated the necessity of the comparator mechanism (e.g. Synofzik *et al.* 2008), the question we address reaches farther, asking whether the comparator model, and specifically the matching component, represents a sufficient mechanism that can account for the sense of agency.

A clear description or definition of a sense of agency is required to assess the sufficiency of the comparator model. Although various slightly different definitions have been used in the literature [e.g. the feeling that I cause my actions (Gallagher 2000), the feeling of intending and executing my actions (Tsakiris and Haggard 2005), the feeling that I cause events through my actions (Haggard and Chambon 2012)], a minimal definition that fits the comparator model as used in experimental paradigms defines the sense of agency as a result of individual actions and their direct consequences, namely *the feeling that my action caused an event in the outside world* (Haggard 2005). Within the scope of this article, we use this definition of a sense of agency because the comparator mechanism functions by virtue of a motor signal (generating action *a*) and the subsequent sensory consequences (the occurrence and detection of the predicted event *e*), which is then assumed to lead to a sense of agency. If the comparator model is sufficient for any type of sense of agency, it would be for the one corresponding to this definition (see also Wong 2012).

Following from this definition, several subcomponents of a sense of agency can be identified. Namely, the sense of agency requires an internal representation of an action by the agent (an ownership predicate), a perceived event and an inferred causal relation between the two. Additionally, to account for the corresponding feeling, a phenomenological dimension is required. We will consider the latter beyond the scope of this article and for the sake of our argument assume that an organism capable experiencing a sense of agency must have the internally represented content such that it presents an event as being caused by a self-produced action (e.g. Crane 2003; Chalmers 2004; Bayne and Levy 2006). In the case of artificial systems, the question of the possibility of phenomenological experiences remain unclear on a fundamental level and will require significant theoretical work before becoming meaningfully addressable (but see Zaadnoordijk and Besold 2018, for a discussion on functionally equivalent implementations of phenomenological states). Also note that leaving the phenomenological dimension aside does not reduce the generality of the presented argument, which rests exclusively on the representational capacity of the matching component of the comparator model. Thus, our reasoning holds notwithstanding the presence or absence of phenomenological considerations.

As outlined in the Introduction section, researchers have focused on participants' capacity to detect the match between the predicted sensory consequences of their actions and their observed sensory consequences. In order for this to yield a sense of agency without additional steps, the match itself would have to represent the content “my action a caused event e.” The representation of the content “my action a caused event e” presupposes the representation of causal relations (e.g. cause(a, e)), ownership predicates (e.g. mine(a)), and relations between them. However, the nature of a match is to merely code for an equality relation: equal(prediction, observation). Detecting a match may therefore at best result in the content of “the sensory prediction is equal to the sensory observation.” However, the representation of the equality relation does not express the causal relations, nor the ownership predicates, nor the complexity of the interaction between these, required for a sense of agency; it remains fully agnostic to all these dimensions as it expresses only and exclusively the equality relation. A subsequent process is required for the agent to go from the representation of the equality to a causal representation. The latter thus cannot be accounted for by the matching process alone. A second scenario that is even further removed from generating a sense of agency is conceivable too: the match detection mechanism may not even lead to a representation of the equality relation. After all, it may be a low-level categorization mechanism that classifies signals as belonging to distinct classes of events without generating a representation of those events as “sensory prediction (not) equal to sensory observation.” If this is the case, the mechanism does not yield any representation about the content of the signal, it just detects two types of signals that are different based on the presence or absence of equality between prediction and observation but, unlike the previous scenario, without making such a representation accessible to the agent, and categorizes them into two different “bins.” Again, one or more subsequent processes have to be introduced into the model, closing the gap from low-level categorization to the ultimately required causal interpretation. The need for these subsequent processes demonstrates that the match detection, on which researchers thus far have focused, is by itself not sufficient to be taken as evidence for the capacity for a sense of agency. It may therefore well be that an agent’s comparator mechanism functions perfectly, but without the additional inference a sense of agency will still not arise.

Since the comparator model does seem to successfully categorize self-produced and externally produced sensory inputs, can it then be regarded as a model for self-other distinction? Unfortunately, we again argue that it cannot. The system may have the capacity to classify the signals into two bins, but it cannot generate the labels for the bins. Either one has to assume that the labels have been obtained at an earlier stage or that the labels are concurrently inferred through some other process. It is possible that the match is taken as input to an inferential process (as well as any additional cues such as proprioceptive signals), but this requires assuming another mechanism in addition to the matching mechanism, namely one that carries out the inferential step. An action may produce additional cues, such as proprioceptive signals, and the inferential process may indeed be facilitated by the presence of these cues. However, next to being a potentially irrelevant cue (e.g. one might move and receive proprioceptive signals even when the sensory input was not caused by them), the availability of these cues does not resolve the need for an inference. This invalidates the sufficiency of the matching mechanism for self-other distinction as such.

Some readers might argue that according to some accounts of the comparator model, it is not just the match that produces the sense of agency but the match combined with additional processes. We grant that the “match-only version” may not have been how the model was intended from a theoretical point of view and that within the comparator model proper the match is not considered sufficient to generate a sense of agency. However, we counter that this is how the comparator model has often been used and interpreted in certain lines of research. This becomes, e.g., evident in artificial implementations of the sense of agency when a system that merely learned to compare its sensory predictions and observations is said to have a sense of agency (Pitti *et al.* 2009; Brody 2016). Moreover, experiments have been set up according to this interpretation of the model: they tap into the ability to detect the congruence of sensorimotor contingencies but fail to test for the ability to make the subsequent inference that when a match is detected, the action is likely to be caused by oneself (see e.g. David *et al.* 2008, for an extensive literature review on the use of the comparator model in sense of agency experiments in adults). Developmental scientists have designed the infant equivalent of these contingency-detection experiments to test for the presence of a sense of agency or body-awareness (Schmuckler 1996; Gergely and Watson 1999; Rochat and Striano 2000; Schmuckler and Jewel 2007; Watanabe and Taga 2011). When testing adults’ sense of agency, researchers tend to take this inferential process for granted, but developmental psychologists and roboticists cannot afford the same luxury.

In sum, while the comparator model might be valuable in its theoretical formulation, our analysis shows that what we have been considering in theoretical discussions, as well as what we have been testing in practice, may not be the mechanism that produces the sense of agency but rather a condition under which the sense of agency is experienced if, and only if, all other requisite processes have been fully developed or, in case of artificial systems, have been implemented.

Grounding Theory in Practice

In this section, we will ground the aforementioned considerations by demonstrating their consequences in an example from the domain of developmental robotics. The benefit of robotics experiments is that we know what has been implemented and thus do not have to guess about the available modules (and corresponding mechanisms) underlying the agent’s capacities. This makes them more suitable for illustrative purposes than infant studies, although it is important to note that the same line of arguments holds for infant studies as well. As mentioned earlier, making a first step towards building the developing sense of agency in artificial agents, Schillaci *et al.* (2016) set up a robotics study in which they made use of forward models to distinguish between self-produced and externally produced signals. In their study, the robot learned to respond differently to signals that it could predict (self-produced) by attenuating them compared to unpredictable signals (externally produced). Providing insight into their reasoning of how these results relate to a developing sense of agency, the authors state: “[…] our experiment shows that prediction errors generated by sensorimotor simulations are smaller when the proprioceptive and motor information are coherent with the perceived ego-noise. Simply put, sensory attenuation is more pronounced when the robot is the owner of the action. When this is not the case, sensory attenuation is worse, as the incongruence of the proprioceptive and motor information with the perceived ego-noise generates bigger prediction errors, which may constitute an element of surprise for the agent and allow it to distinguish between self-generated actions and those generated by other individuals.” (p. 396) As described before this conception of a sense of agency is in line with previous work, but it leaves the open question: how does the robot know the signals were produced by itself? It may objectively be true that the signals were self-generated, but this is only relevant in the context of a sense of agency and self-other distinction if the robot is able to learn that a match means that signals are self-generated. (In reality, other contextual factors likely disallow the match and the self-generated signals to have a one-to-one relation. For the sake of simplicity, we do not take these into account here and only briefly mention them in the discussion section.) As we see it, there are three possible scenarios:

1. The robot does not know. It merely categorizes the two signals by virtue of similarity on signal level between the incoming signal and the predicted signal and the researchers interpret the two categories as "self" and "other."
2. The labels "self" and "other" were hard-coded into the robot by the researchers. That is, the robot categorizes the signals as in scenario 1, and assigns labels based on pre-provided criteria.
3. The robot uses the match to make an inference about how that match could have come about (namely, that the sensory input was caused by its own earlier actions) and that the robot itself is a distinctive entity in the world.

In the first two scenarios, the robot merely categorizes the signals into two separate bins to select the signal it needs to attenuate. The sensory attenuation process is impressive from an engineering perspective but not as much from a cognitive point of view. There are many systems that are comparable to the sensory attenuation task in that they can perform categorization tasks (e.g. Leemans *et al.* 2002; Nawrocky *et al.* 2010) or use properties of incoming signals as the basis for their next actions (e.g. Bahdanau *et al.* 2014; Yeh *et al.* 2017). Still, these systems work on a purely statistical basis, relying on elaborate forms of pattern matching, without requiring an understanding of the concepts they are working with. Similarly, the ability to categorize self and externally produced signals is not sufficient for the development of concepts such as "self" and "other." Our argument is similar in structure and implications for concept learning to the core point made by Mandler (1988, p. 117), who wrote: "We should not be misled by the complexity of these perceptual processing mechanisms. They are sophisticated, of course, but then so are the perceptual processing mechanisms of most organisms or, for that matter, the industrial vision machines that neatly discriminate nuts from bolts. To categorize incoming stimuli into different types is a basic component of a perceptual recognition device; by itself, this ability tells us nothing about the formation of accessible concepts that may be used for purposes of thought and reflection. The industrial machine may throw nuts into one bin and bolts into another (making its choices by, e.g., computing the ratio of the diameter of each object to its perimeter), but we would not want to say that it has a concept of nuts and bolts." One may now argue, as described in scenario 2, that the labels may be provided to the robot at this stage in its development, because equivalently infants may have obtained the labels elsewhere during an earlier stage in development. In this case, the "self" and "other" labels can be assumed to be part of their knowledge, making it unproblematic for the researchers to provide the robot with hard-coded labels as well. Nevertheless, even if the labels are provided, the match will only bring the agent as far as to categorize the signals into two bins (similar to scenario 1) but now with labels on these bins. As categorization does not automatically lead to a causal inference, the match itself will still not generate a causal representation of the consequences of the agent's actions required for a sense of agency.

Only in the third scenario would the robot perform an intelligent and cognitively demanding task that could lead to a sense of agency. However, in this case it is in fact not the match that generated the sense of agency and self-other distinction but the subsequent inference made based on that match. The additional inferential process is essential to generating the sense of agency. Hence, in all three scenarios the match itself in isolation is insufficient to produce a sense of agency. Thus, although being a sensible first step, focusing exclusively on the capacity to detect a match will not lead to conclusive insights about the emergence of a sense of agency in developmental studies, or to successful engineering of a sense of agency in robotics.

Future Directions

In the sections above, we have argued that testing for the presence of a match does not inform the researcher about the presence of the causal action-effect representation required for a sense of agency. The causal relation between the action and the subsequent sensory effect seems to be inferred by adult participants after detecting a match. This means that in our conceptualization of the comparator model, this inferential process must be added to reflect the actual processing (see Fig. 2). Making the inferential step explicit in the model is important to show which processes must be tested for in developmental science and implemented in artificial intelligence.

Designing an experiment to assess the presence of a sense of agency, for instance in developmental science, thus means that researchers have to find evidence for the ability to detect the match (for instance by demonstrating infants' sensorimotor contingency detection) as well as evidence for the subsequent causal interpretation of that match. The presence of such a causal model could be tested with measures of anticipatory behavior (see e.g. Kenward 2010; Miyazaki *et al.* 2014) or measures of a violation of expectation. Infants' prediction that their action will be followed by an effect yields a strong indication that they have built an action-effect model. Testing for infants' causal models requires research into what type of behavioral or physiological markers indicate the presence of such a model. Computer simulations of infant behavior in the aforementioned mobile-paradigm recently showed that the previously reported increased movement behavior during the connect phase (when the ribbon is connected to the limb and the sensorimotor contingency is active) cannot be taken as evidence for an underlying causal action-effect model as the behavioral pattern could be explained by a simpler cognitive mechanism (Zaadnoordijk *et al.* 2018). These simulations pointed towards the disconnect phase (when the sensorimotor contingency ceases to exist) as the phase that potentially is better able to distinguish between those infants who have learned the causal relation and those who have not. Based on the computer simulation work, Zaadnoordijk *et al.* (submitted) hypothesized that if infants had learned the causal relation between their action and the mobile movement, they would show indicators of a violation of expectation upon cessation of the sensorimotor contingency. The results of their electroencephalography (EEG) study with 3- to 4.5-month olds show that indeed that the group of infants who showed a neural violation of expectation also showed an extinction burst (i.e. an additional increase and then decrease of movement frequency relative to when the contingency was still active) in their behavior. Both these measures suggest that the infant has a prediction error about the causal action-effect relation, which they are trying to resolve with additional movements. More experimental research is required to fully

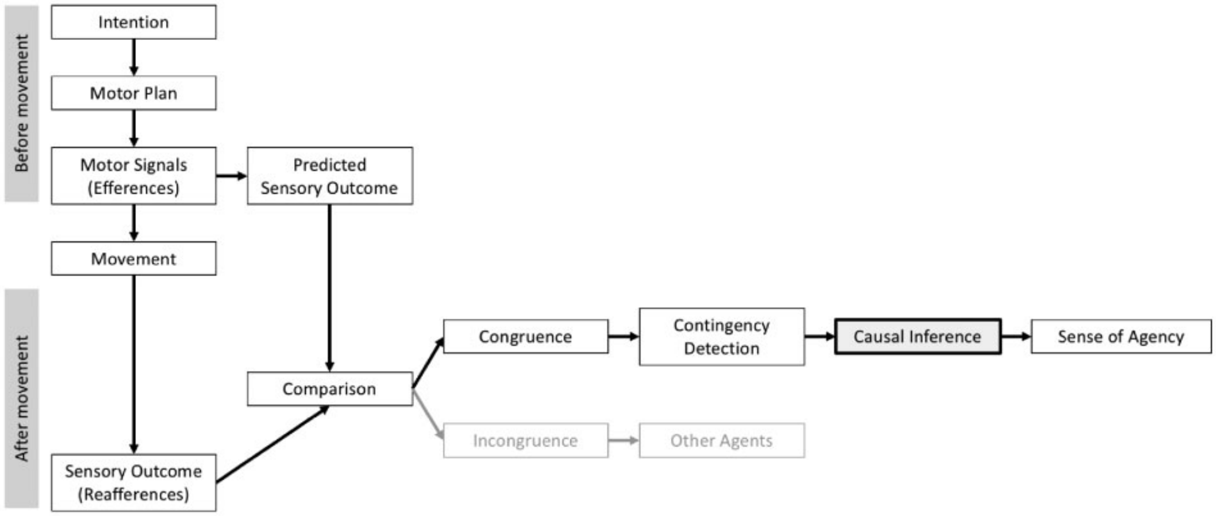

Figure 2. An updated schematic overview of the comparator model. (In this figure we do not address the case of incongruence between the predicted and observed sensory outcomes, although it is likelythat additional inferential processes take place after a mismatch to adequately attribute the sensory information to a source. Moreover, the figure only touches upon the predictive components of thesense of agency and does not consider any additional processes that might be necessary or sufficient for the judgment of agency. Previous work on this topic has been carried out by, e.g., Synofzik et al. (2008, 2013).)

understand how these models are built and what their exact relation is to the feeling of agency. Moreover, ideally, one would additionally demonstrate that infants show a violation of expectation if the effect occurs in absence of any action as well, but this form of causal learning may be too advanced in early infancy. The mentioned measures have already been used in other domains, making it relatively easy to apply them to developmental research into the sense of agency. This small change to experimental designs would greatly increase our understanding of the developing sense of agency in infancy.

For artificial intelligence, the consequences of an extension of the comparator model entail changes in the implemented cognitive architecture (or the learning goal) of the agent. In the extended model mere contingency detection is not sufficient, but instead the capacity to successfully perform a causal inference is required to confirm that the observed change in the world state has indeed been caused by the agent as independent actor in its environment. This in turn necessitates several preconditions to be met, among others requiring the agent (i) to have a model of the environment and of itself (so as to also be able to perform a form of self-other distinction), (ii) to be able to re-represent the information regarding the change in the environment and the detected match in a form usable for causal reasoning, and (iii) to be equipped with the required reasoning capabilities to actually perform the corresponding causal inferences. This ties into several long-standing research objectives in artificial intelligence, including both high-level perception (e.g. Chalmers *et al.* 1992) to bridge from the contingency detection to the representations required for the causal inference, and causal modeling and reasoning (e.g. Shoham 1988; Pearl 2009) to perform the actual inference step.

It is, as of yet, not clear how agents develop from detecting a sensorimotor contingency to inferring a causal relation between their actions and the sensory effects, nor which components constitute the sense of agency. We have made a case for the necessity of a causal representation, but arguably the presence of a causal representation alone does not generate a sense of agency. At the very least, as mentioned before, an ownership predicate for the action will be needed, and for the phenomenological aspect there must be the capacity for phenomenological experiences. However, this list is likely incomplete and more (theoretical and experimental) research is needed to come to a full understanding of the relevant components and processes to the experience of agency. The corresponding open questions will be crucial to answer for both developmental science and artificial intelligence in the context of the developing sense of agency. They span various active areas of work receiving significant attention from the respective scientific communities. Advances made in any of these domains will play an important role in accomplishing the difficult task of imbuing robots with a sense of agency.

Discussion

Applying models obtained through experiments with adults to research into the development of capacities in infants or artificial systems may have important caveats. Since cognitive capacities in adults have manifested in a full-fledged manner, experimental research tends to focus on the conditions under which the capacity is active or modulated. This research investigates a particular necessary condition or process associated to capacities that have already fully developed and can thus ignore other necessary processes as they can be assumed to occur. In developmental science and artificial intelligence, on the other hand, the research questions often pertain to what components constitute the capacity and whether each of those components are present (either through development or by implementation). These investigations target the sufficiency of a set of processes and mechanisms for the cognitive capacity. Directly applying models and experimental paradigms used in adult research to infant research or artificial intelligence can therefore lead to answering questions about a necessary process while a sufficient process had been the intended topic of investigation.

In this article, we showcased such a caveat in the context of the comparator model as a model for the sense of agency. Following experiments from the field of cognitive (neuro)science with adult participants, developmental psychologists have turned to the comparator model as a starting point for research into the development of the sense of agency. Against that backdrop, we presented a conceptual analysis of the explanatory value of the model for the sense of agency. In doing so, we focused specifically on the match detection mechanism as subcomponent of the comparator model tested in experiments with adults and infants and built into artificial agents as a means for them to develop a sense of agency. Our conceptual analysis shows that the match detection mechanism postulated by the comparator model is insufficient to produce a sense of agency, as it lacks the capacity to represent causal relations, ownership predicates and the complex interaction between them. Although the model may describe a mechanism that plays a role in the sense of agency, we demonstrated that what is being tested in practice may not be the mechanism that can explain, and thus produce, the sense of agency. This has far-reaching consequences for current experimental research practice: As the mechanism lacks the capacity to produce a sense of agency, experiments focusing on the development of this mechanism alone cannot lead to conclusions about the emergence or presence of a sense of agency. Moreover, it means that one or more additional modules are required when engineering a sense of agency into an artificial agent.

This is not the first article that addresses the suitability of the comparator model in the context of the sense of agency. Another criticism of the comparator model has targeted the assumption that the feeling of agency underlies the judgment of agency that is often measured in experimental settings (Grünbaum 2015). Grünbaum reasons that there is no need to assume that the comparator mechanism, as low-level motor mechanism, generates the feeling of agency, because the data in judgment of agency experiments can be explained equally well without ascribing this phenomenological quality to the participants. He argues that since the judgment of agency is partially generated by the comparator mechanism and can be explained without the feeling of agency, the hypothesis that the comparator model also generates the feeling of agency is not parsimonious. Recently, Christensen and Grünbaum (2018) further questioned the explanatory value of the comparator model by introducing a differentiation between broad and narrow sense of agency. The authors argue that in the field researchers typically treat the sense of agency both as associated to the voluntariness of movements (narrow sense of agency) as well as to the consequences of those movements in the external world (broad sense of agency). Conflating the broad and narrow notion is problematic because the comparator model, if anything, would be a model for the narrow sense of agency, but the experimental practice tends to focus on the sensory consequences of an action, which the comparator model is unlikely to be able to compute (Christensen and Grünbaum 2018). Our argument differs from Grünbaum's argument in several ways. Since our article is partly targeting infant researchers, we are not as concerned with the relation between the feeling of agency and the judgment of agency, as infants are not yet capable of the latter. Instead, we aimed to assess whether the comparator model would in principle be able to generate the feeling of agency (which we refer to as the sense of agency in our paper). Unlike Grünbaum, we do not conclude that the feeling of agency might not exist as such, but that the comparison between the predicted and observed sensory information is not sufficient to generate the phenomenal experience. Since our argument asserts that the match in the comparator model cannot generate any sense of agency, the distinction between the broad and narrow sense of agency is not decisive for our claims. However, the relevance of the differentiation is clear from a mechanistic point of view and any research into explanatory models of the sense of agency ought to take it into account when formulating the underlying processes and mechanisms.

Closer to our concern are the criticisms uttered by Synofzik *et al.* (2008). After a thorough analysis of the empirical literature, they argued that the comparator model lacks a conceptual division between the feeling of agency and the judgment of agency. These authors consider the feeling of agency to be a low-level non-conceptual feeling of being an agent, which comes about through a predictive process. The judgment of agency on the other hand refers to the conclusion that one is an agent following a postdictive reasoning process (see also Synofzik *et al.* 2013). The authors rightly pointed out that the capacity tested in experiments sometimes pertains to the feeling of agency (e.g. Lindner *et al.* 2005) and sometimes to the judgment of agency (e.g. Aarts *et al.* 2005). In their conceptual analysis, they then went on to argue that the comparator model in fact cannot fully explain either of these phenomena. Their argument is based on two types of evidence. First, based on the literature they observed that participants may self-attribute actions even when there is a mismatch between the predicted and actual consequence, and that therefore a mismatch is not sufficient to prevent the experience of agency. Second, Synofzik *et al.* (2008) argued that a patient study suggested that comparing the efference and reafference signals is not always sufficient for a sense of agency. Based on this evidence, they conclude that there are cases for which the comparator output is neither a sufficient nor a necessary condition for the feeling of agency nor the judgment of agency. A related concern regarding its sufficiency and necessity has been raised by Mylopoulos (2017) who argued that the sense of agency may be considered the default state and that only when the sense of agency is violated, one becomes aware of one's agency. Since the default state is unrelated to motor activity, Mylopoulos suggests abandoning the comparator model as explanation for the sense of agency altogether. We do not believe that the sense of agency is the default state, because there can be a state of no sense of agency also in the absence of a violation of the sense of agency. But even if one assumes that the sense of agency is a default state, from a developmental and robotics point of view, it is not clear how the acquisition of and development into that default state would occur and whether the comparator model could play a part in this process.

An important difference between the criticism of Synofzik *et al.* (2008) and the criticism presented here is that our argument holds more generally. Our analysis additionally covers those cases, which had previously not been accounted for, in which the match between the predicted and the actual consequence leads to a sense of agency in typically developed adults (i.e. what can be taken as the most frequently occurring "standard case" of sense of agency). As explained above, notwithstanding the existence of these cases, for principled reasons the match detection mechanism does not explain or produce the phenomenon. In this sense, standing on firm theoretical ground, we push the insufficiency argument further than Synofzik *et al.* (2008). While their theory still allows for the comparator model's match to lead to a sense of agency in certain situations by itself, we argue that this can never be the case without additional inferences. Hence, when researchers wish to assess the presence or emergence of a sense of agency, they cannot base their conclusions solely on the match detection but need to also determine whether the subsequent inference is made. Evidence for the subsequent inferential process would demonstrate that the capacity to make this type of inference has emerged and can readily be used.

The importance of this difference lies in the consequences it has for interpreting data from other populations, such as infants and robots, in which a developed sense of agency cannot be assumed *a priori*. As it has theoretically been shown that detecting the match is in itself not sufficient to infer that events are caused by one's own actions, providing evidence that infants can detect the match is therefore insufficient to conclude that they are at the same time able to use the match detection to infer that they were the cause of an effect and, thus, to experience a sense of agency. The insufficiency of the match detection for a sense of agency naturally also holds for the adult population. However, the problematic nature of it usually remains unnoticed because in this population researchers (often without explicit mention) assume that participants are capable of making this inference and experiencing their agency. Consequently, researchers do not attempt to find evidence of the inferential ability; they are instead focused on the conditions under which the subsequent experience of agency can be altered (e.g. by violating the sensorimotor contingency).

In this article, we left untouched several other concerns regarding the comparator model as a model of the sense of agency. For instance, it remains unclear what the nature of a match between the predicted and the observed sensory consequences entails, what the relevant aspects and levels of detail are on which the match is made, and whether different degrees to which a sense of agency is experienced and the observed degrees of sensory attenuation in the brain can be captured by a seemingly binary match/no-match categorization between predicted and observed sensory consequences. These are questions that so far have received only little attention in the context of a sense of agency, especially in artificial agents. Some of these questions, e.g. about the binary nature of the match and its implications, may become obsolete as researchers further incorporate the predictive processing accounts (Friston 2010; Hohwy 2007; Apps and Tsakiris 2014) into this framework, shifting the focus from match to degree of prediction error. Predictive processing allows for a flexible processing and integration of cues and more naturally accommodates for the possibility of graded experiences than the comparator model as currently used in the context of a sense of agency. The degree of prediction error could align to the strength of sensory attenuation, better capturing situations in which agents are unsure about whether or not they were the cause of a sensory effect. The degree of prediction error can be modulated by many different cues that are weighted differently in various contexts. Nevertheless, an implementation or even a computational-level theory of causal inference is still in early stages (see e.g. Gopnik *et al.* 2004; Otworowska *et al.* 2016 for related work), also in predictive processing theories, and tractability problems related to causal learning have not yet been resolved. This issue as well as other questions, such as the relevant aspects that are being compared and the levels at which the comparison is made, remain relevant and need to be addressed in future theoretical work. Moreover, the role that environmental and situational context plays in the generation of the predictions is poorly understood. Whether a match leads to a sense of agency seems to be top-down penetrable by world knowledge. This matter was briefly touched upon by Zaadnoordijk *et al.* (2015) but requires a more thorough analysis.

Since the conceptual analysis of the comparator model has shown that the match detection mechanism lacks the required representational capacity, this opens up the question of what type of mechanism can account for these representations. Although an attempt to answer this question would be outside the scope of this article, we think that there is promising research into this direction worth mentioning. For instance, Nagai and Asada (2015) as well as Otworowska *et al.* (2016) have implemented predictive processing accounts into artificial agents to learn causal relations between their actions and the consequences. While these implementations are not yet able to fully account for the cognitive capacity, they have the potential to elucidate the type of mechanisms required for causal representations. Closing the existing gaps in the explanatory model will bring full understanding of the developing (biological and artificial) sense of agency within reach.

Acknowledgments

The authors thank Iris van Rooij for her valuable contributions to this research. Parts of this research have been carried out at City, University of London. This research was supported by an Aspasia Prize of the Netherlands Organisation for Scientific Research (NWO) awarded to SH.

Conflict of interest statement. None declared.

References

Aarts H, Custers R, Wegner DM. On the inference of personal authorship: enhancing experienced agency by priming effect information. *Conscious Cogn* 2005;**14**:439–58.

Apps MA, Tsakiris M. The free-energy self: A predictive coding account of self-recognition. *Neurosci Biobehav Rev* 2014;**41**: 85–97.

Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and *translate*. *arXiv Preprint arXiv:1409.0473*, 2014.

Bahrick LE, Watson JS. Detection of intermodal proprioceptive–visual contingency as a potential basis of self-perception in infancy. *Dev Psychol* 1985;**21**:963.

Banks MS, Salapatek P. Infant pattern vision: a new approach based on the contrast sensitivity function. *J Exp Child Psychol* 1981;**31**:1–45.

Bayne T, Levy N. The feeling of doing: deconstructing the phenomenology of agency. In: Sebanz N and Prinz W (eds), *Disorders of Volition*. Cambridge, MA: MIT Press, 2006, 49–68.

Bechtel W. Looking down, around, and up: mechanistic explanation in psychology. *Philos Psychol* 2009;**22**:543–64.

Bechtel W, Abrahamsen A. Explanation: a mechanist alternative. *Stud Hist Philos Sci C* 2005;**36**:421–41.

Bechtle S, Schillaci G, Hafner VV. On the sense of agency and of object permanence in robots. In: *Proceedings of the Joint IEEE International Conference on Development and Learning and Epigenetic Robotics (ICDL-EpiRob)*. Cergy-Pontoise: IEEE, 2016, 166–71.

Blakemore SJ, Frith CD, Wolpert DM. Spatio-temporal prediction modulates the perception of self-produced stimuli. *J Cogn Neurosci* 1999;**11**:551–9.

Blakemore SJ, Wolpert DM, Frith CD. Abnormalities in the awareness of action. *Trends Cogn Sci* 2002;**6**:237–42.

Brody J. The enacted KOAN—an agent's knowledge of agency. *Procedia Comput Sci* 2016;**88**:211–6.

Burghart C, Mikut R, Stiefelhagen R, *et al.* A cognitive architecture for a humanoid robot: A first approach. In: *Proceedings of the 5th IEEE-RAS International Conference on Humanoid Robots*. Tsukuba: IEEE, 2005, 357–62.

Chalmers D. The representational character of experience. In: Leiter B (ed.), *The Future for Philosophy*. Oxford: Clarendon Press, 2004, 153–81.

Chalmers DJ, French RM, Hofstadter DR. High-level perception, representation, and analogy: a critique of artificial intelligence methodology. *J Exp Theor Artif Intell* 1992;**4**:185–211.

Chambon V, Sidarus N, Haggard P. From action intentions to action effects: how does the sense of agency come about? *Front Hum Neurosci* 2014;**8**:320.

Christensen MS, Grünbaum T. Sense of agency for movements. *Conscious Cogn* 2018;**65**:27–47.

Crane T. The intentional structure of consciousness. In: Jokic A and Smith Q (eds), *Consciousness: New Philosophical Perspectives*. Oxford and New York: Oxford University Press, 2003, 33–56.

David N. New frontiers in the neuroscience of the sense of agency. *Front Hum Neurosci* 2012;**6**:10–3389.

David N, Newen A, Vogeley K. The "sense of agency" and its underlying cognitive and neural mechanisms. *Conscious Cogn* 2008;**17**:523–34.

Farrer C, Franck N, Frith CD, *et al.* Neural correlates of action attribution in schizophrenia. *Psychiatry Res* 2004;**131**:31–44.

Farrer C, Franck N, Georgieff N, *et al.* Modulating the experience of agency: a positron emission tomography study. *Neuroimage* 2003;**18**:324–33.

Fourneret P, Jeannerod M. Limited conscious monitoring of motor performance in normal subjects. *Neuropsychologia* 1998;**36**: 1133–40.

Friston K. The free-energy principle: a unified brain theory? *Nat Rev Neurosci* 2010;**11**:127–38.

Frith C. The self in action: Lessons from delusions of control. *Conscious Cognition* 2005;**14**:752–70.

Frith CD, Blakemore SJ, Wolpert DM. Explaining the symptoms of schizophrenia: abnormalities in the awareness of action. *Brain Res Rev* 2000;**31**:357–63.

Gallagher S. Philosophical conceptions of the self: Implications for cognitive science. *Trends Cogn Sci* 2000;**4**:14–21.

Gergely G, Watson JS. Early socio-emotional development: contingency perception and the social-biofeedback model. In: Rochat P (ed.), *Early Social Cognition: Understanding Others in the First Months of Life*. Mahwah, NJ: Erlbaum, 1999, 101–36.

Gopnik A, Glymour C, Sobel DM, *et al.* A theory of causal learning in children: causal maps and Bayes nets. *Psychol Rev* 2004;**111**:3.

Grünbaum T. The feeling of agency hypothesis: a critique. *Synthese* 2015;**192**:3313–37.

Haggard P. Conscious intention and motor cognition. *Trends Cogn Sci* 2005;**9**:290–5.

Haggard P, Chambon V. Sense of agency. *Curr Biol* 2012;**22**: R390–2.

Heathcock JC, Bhat AN, Lobo MA, *et al.* The performance of infants born preterm and full-term in the mobile paradigm: learning and memory. *Phys Ther* 2004;**84**:808–21.

Heathcock JC, Bhat AN, Lobo MA, *et al.* The relative kicking frequency of infants born full-term and preterm during learning and short-term and long-term memory periods of the mobile paradigm. *Phys Ther* 2005;**85**:8–18.

Hohwy J. The sense of self in the phenomenology of agency and perception. *Psyche* 2007;**13**:1–20.

Jeannerod M. The sense of agency and its disturbances in schizophrenia: a reappraisal. *Exp Brain Res* 2009;**192**:527–32.

Johnson MH. *Developmental Cognitive Neuroscience*. London: Blackwell, 1997.

Kenward B. 10-month-olds visually anticipate an outcome contingent on their own action. *Infancy* 2010;**15**:337–61.

Kelso JAS. On the self-organizing origins of agency. *Trends Cogn Sci* 2016;**20**:490–9.

Lagnado DA, Sloman SA. Learning causal structure. In: Gray W and Schunn CD (eds), *Proceedings of the Twenty-fourth Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Erlbaum, 2002, 560–5.

Leemans V, Magein H, Destain MF. AE—automation and emerging technologies: on-line fruit grading according to their external quality using machine vision. *Biosyst Eng* 2002;**83**:397–404.

Lindner A, Thier P, Kircher TT, *et al.* Disorders of agency in schizophrenia correlate with an inability to compensate for the sensory consequences of actions. *Curr Biol* 2005;**15**: 1119–24.

Mandler JM. How to build a baby: on the development of an accessible representational system. *Cogn Dev* 1988;**3**:113–36.

Miall RC, Wolpert DM. Forward models for physiological motor control. *Neural Netw* 1996;**9**:1265–79.

Miyazaki M, Takahashi H, Rolf M, *et al.* The image-scratch paradigm: a new paradigm for evaluating infants' motivated gaze control. *Sci Rep* 2014;**4**:5498.

Mylopoulos M. A cognitive account of agentive awareness. *Mind Lang* 2017;**32**:545–63.

Nagai Y, Asada M. Predictive learning of sensorimotor information as a key for cognitive development. In: *Proceedings of the IROS 2015 Workshop on Sensorimotor Contingencies for Robotics*. Hamburg: IEEE, 2015.

Nawrocky M, Schuurman DC, Fortuna J. Visual sorting of recyclable goods using a support vector machine. In: *Proceedings of the Canadian Conference on Electrical and Computer Engineering (CCECE)*. Calgary: IEEE, 2010, 1–4.

Otworowska M, Zaadnoordijk L, de Wolff E, *et al.* Causal learning in the crib: a predictive processing formalization and babybot simulation. In: *Proceedings of the Joint IEEE International Conference on Development and Learning and Epigenetic Robotics (ICDL-EpiRob)*. Cergy-Pontoise: IEEE, 2016, 39–40.

Pearl J. *Causality*. Cambridge, MA: Cambridge University Press, 2009.

Pico A, Schillaci G, Hafner VV, *et al.* How do I sound like? Forward models for robot ego-noise prediction. In: *Proceedings of the Joint IEEE International Conference on Development and Learning and Epigenetic Robotics (ICDL-EpiRob)*. Cergy-Pontiose: IEEE, 2016, 246–51.

Pitti A, Mori H, Kouzuma S, *et al.* Contingency perception and agency measure in visuo-motor spiking neural networks. *IEEE Trans Auton Ment Dev* 2009;**1**:86–97.

Qiao H, Wu W, Yin P. Biologically inspired models for visual cognition and motion control: an exploration of brain-inspired intelligent robotics. *Science* 2016;**354**:35–8.

Rochat P, Morgan R. Spatial determinants in the perception of self-produced leg movements in 3-to 5-month-old infants. *Dev Psychol* 1995;**31**:626.

Rochat P, Morgan R. Two functional orientations of self-exploration in infancy. *Br J Dev Psychol* 1998;**16**:139–54.

Rochat P, Striano T. Emerging self-exploration by 2-month-old infants. *Dev Sci* 1999;**2**:206–18.

Rochat P, Striano T. Perceived self in infancy. *Infant Behav Dev* 2000;**23**:513–30.

Rovee CK, Rovee DT. Conjugate reinforcement of infant exploratory behavior. *J Exp Child Psychol* 1969;**8**:33–9.

Rovee-Collier CK, Morrongiello BA, Aron M, *et al.* Topographical response differentiation and reversal in 3-month-old infants. *Infant Behav Dev* 1978;**1**:323–33.

Sabes PN. The planning and control of reaching movements. *Curr Opin Neurobiol* 2000;**10**:740–6.

Sato A, Yasuda A. Illusion of sense of self-agency: discrepancy between the predicted and actual sensory consequences of actions modulates the sense of self-agency, but not the sense of self-ownership. *Cognition* 2005;**94**:241–55.

Schillaci G, Ritter CN, Hafner VV, *et al*. Body representations for robot ego-noise modelling and prediction. Towards the development of a sense of agency in artificial agents. In: *Proceedings of the Artificial Life Conference*. Cambridge, MA: MIT Press, 2016, 390–97.

Schmuckler MA. Visual-proprioceptive intermodal perception in infancy. *Infant Behav Dev* 1996;**19**:221–32.

Schmuckler MA, Jewell DT. Infants' visual-proprioceptive intermodal perception with imperfect contingency information. *Dev Psychobiol* 2007;**49**:387–98.

Shadmehr R, Smith MA, Krakauer JW. Error correction, sensory prediction, and adaptation in motor control. *Annu Rev Neurosci* 2010;**33**:89–108.

Shanahan M. A cognitive architecture that combines internal simulation with a global workspace. *Conscious Cogn* 2006;**15**: 422–49.

Shoham Y. *Reasoning About Change: Time and Causation from the Standpoint of Artificial Intelligence*. Cambridge, MA: MIT Press, 1988.

Sperduti M, Delaveau P, Fossati P, *et al*. Different brain structures related to self- and external-agency attribution: a brief review and meta-analysis. *Brain Struct Funct* 2011;**216**:151–7.

Sperry RW. Neural basis of the spontaneous optokinetic response produced by visual inversion. *J Comp Physiol Psychol* 1950;**43**:482.

Synofzik M, Vosgerau G, Newen A. Beyond the comparator model: a multifactorial two-step account of agency. *Conscious Cogn* 2008;**17**:219–39.

Synofzik M, Vosgerau G, Voss M. The experience of agency: an interplay between prediction and postdiction. *Front Psychol* 2013;**4**:127.

Tsakiris M, Haggard P. Experimenting with the acting self. *Cogn Neuropsychol* 2005;**22**:387–407.

Tsakiris M, Haggard P, Franck N, *et al*. A specific role for efferent information in self-recognition. *Cognition* 2005;**96**:215–31.

Von Holst E, Mittelstaedt H. Das Reafferenzprinzip. *Naturwissenschaften* 1950;**37**:464–76.

Watanabe H, Taga G. General to specific development of movement patterns and memory for contingency between actions and events in young infants. *Infant Behav Dev* 2006;**29**:402–22.

Watanabe H, Taga G. Initial-state dependency of learning in young infants. *Hum Mov Sci* 2011;**30**:125–42.

Watson JS. Smiling, cooing, and "the game". *Merrill-Palmer Q Behav Dev* 1972;**18**:323–39.

Watson JS. Contingency experience in behavioral development. In Immelmann K, Barlow G.W., Petrinovich L., and Main M. (eds), *Behavioral Development: The Bielefeld Interdisciplinary Project*. New York: Cambridge University Press, 1981, 83–9.

Wolpert DM, Flanagan JR. Motor prediction. *Curr Biol* 2001;**11**: R729–32.

Wolpert DM, Kawato M. Multiple paired forward and inverse models for motor control. *Neural Netw* 1998;**11**:1317–29.

Wong HY. A measure of my agency? *Conscious Cogn* 2012;**21**: 48–51.

Yeh RA, Chen C, Lim TY, *et al*. Semantic image inpainting with deep generative models. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu, HI: IEEE, 2017, 5485–93.

Zaadnoordijk L, Besold TR. Artificial phenomenology for human-level artificial intelligence. In: Chella A, Gamez D, Lincoln P, Manzotti R and Pfautz J (eds), *Proceedings of the AAAI Spring Symposium 2019: Towards Conscious AI Systems*. Vol. 2287, paper 24. Stanford, CA: CEUR Workshop Proceedings, 2018.

Zaadnoordijk L, Hunnius S, Meyer M, *et al*. The developing sense of agency: implications from cognitive phenomenology. In: *Proceedings of the Joint IEEE International Conference on Development and Learning and Epigenetic Robotics (ICDL-EpiRob)*. Providence, RI: IEEE, 2015, 114–5.

Zaadnoordijk L, Otworowska M, Kwisthout J, *et al*. Can infants' sense of agency be found in their behavior? Insights from babybot simulations of the mobile-paradigm. *Cognition* 2018; **181**:58–64.