Synthese
DOI 10.1007/s11229-015-0812-3

S.I.: COGNITION

# What are cognitive processes? An example-based approach

**Albert Newen**[1]

Received: 5 June 2014 / Accepted: 24 June 2015


**Abstract** The question “What are cognitive processes?” can be understood variously as meaning “What is the nature of cognitive processes?”, “Can we distinguish epistemically cognitive processes from physical and biochemical processes on the one hand, and from mental or conscious processes on the other?”, and “Can we establish a fruitful notion of cognitive process?” The present aim is to deliver a positive answer to the last question by developing criteria for what would count as a paradigmatic exemplar of a cognitive process, and then to offer the comparator (or feedforward) mechanism as a convincing paradigmatic example. Thus, the paper argues, given the current state of science, we can indeed establish a fruitful scientific notion of a cognitive process. Nevertheless, it is left open whether the example-based characterization ends up as merely highlighting a fruitful convention within the early-twentyfirst century interdisciplinary investigation of intelligent behaviour in humans, animals, and robots, or whether the examples determine a natural kind or a property cluster.

**Keywords** Cognitive processes · Comparator model · Feedforward model · Efference copy · Cognition

## 1 Introduction

Recent philosophy of mind has seen intense discussion concerning whether cognition is brain-bound (Adams and Aizawa 2008) or extended (Clark and Chalmers 1998; Menary 2010), embodied (Gallagher 2005), embedded (Robbins and Aydede 2009) or

✉ Albert Newen
albert.newen@rub.de

[1] Institute of Philosophy II, Ruhr-Universität Bochum, Universitätsstraße 150a, 44780 Bochum, Germany

Published online: 15 July 2015

enacted (Hurley 1998; Noe 2004). These debates presuppose that we are possessed of a sufficiently clear notion of cognition in the first place. It would not be satisfying to cite our folk-psychological notion of cognition as our preferred notion of cognition, since this would divert the debate into an argument about the role of intuitions and common usage in characterizing concepts in the philosophy of mind. Thus, an important background problem for these debates is to establish a sufficiently clear *scientific notion of cognition*; and this problem centrally concerns the characterization of cognitive *processes*, because it is to such processes that the positions just cited specifically appeal.

The question "What are cognitive processes?" can thus be understood as a challenge to answer the more elaborated question "What is the nature of cognitive processes and can we establish a fruitful scientific notion of a cognitive process?" This question divides further. Concerning their nature, we may ask: Are cognitive processes natural kinds, or individuated conventionally? Are we able to characterize cognitive processes in such a way that they can be *ontologically* distinguished from physical and biochemical processes? Or do cognitive processes simply supervene on the latter? And another way to pose the central question would be: Is there a principled *epistemic* way to define cognitive processes which enables us to show that this way of characterizing cognitive processes is explanatorily fruitful in science, or can we justify a definition only pragmatically, by agreeing on some way of talking about certain phenomena which are not only ontologically but even epistemologically reducible to physical and biochemical processes? Relatedly, we may ask: Can the best science of flexible behaviour work without presupposing cognitive processes?

How can we make progress with these questions, given the widely accepted methodological constraints in this field? It is a commonplace nowadays that there cannot be a fruitful analysis which is purely conceptual: any proffered conceptual analysis can only be useful if it is systematically integrated with relevant empirical knowledge, since semantic knowledge is interwoven with world knowledge, and it is a dogmatic dream that we can keep them apart (Quine 1951). Another presupposition of my investigation here is the acceptance of a naturalistic account of the mind which subscribes to ontological reductionism; yet it is also a commonplace that this leaves open the question of how the epistemological relation between cognitive and biochemical (or physical) processes should best be described. How, indeed, should we begin? Should we first try to define either the cognitive sciences or the cognitive systems, and on that basis try to characterize cognitive processes; or can we approach cognitive processes directly?

Let us have a short look at approaches which aim to *define* cognitive science, developing arguments to the effect that it is doubtful that three of them can furnish a successful strategy, while the fourth, the multiple criteria view, can indeed be developed into a viable account. The four approaches in question are: (i) *Phenomena-based* accounts, according to which cognitive science is defined by the kinds of phenomena that are under investigation, e.g., memory, reasoning, planning, language usage, and problem solving. However, it is unlikely that such accounts will provide the basis of an account of cognitive processes, since these kinds of phenomena are also studied in molecular biology or in the medical sciences, where they are usually not seen as being part of the disciplines belonging specifically to the cognitive sciences (Miller 2003). (ii) *Theory-based* accounts, according to which cognitive science is defined by shared theoretical presuppositions about how intelligent behavior should be explained, namely that it should be explained by reference to mental representation and computation (Thagard 2010; Sect. 4). But even if this was true for the cognitive sciences until the 1990s, these commitments are no longer universally accepted. The cognitive sciences now include a strong group of researchers who base their work on either skepticism concerning or an explicit denial of mental representation and computation (Brooks 1991)—and of course it would be indefensible to exclude nonrepresentational explanations of intelligent behavior in robots, or humans, from cognitive science. (iii) *Community-based* accounts: should we, then, simply say that cognitive science is what researchers who call themselves "cognitive scientists" do? This would mean abandoning hope for any systematic understanding of research projects within cognitive science. In response to these problems, I propose that we can make some progress if we refuse to focus only on one criterion but instead combine several criteria, leading to (iv) a *multiple criteria* account: a research project is part of the cognitive sciences if it deals with one and the same research question concerning cognitive (i.e. minimally flexible or intelligent) behaviour as anchored in the cognitive processes of cognitive systems using different respective cognitive methodologies. This looks at first glance like a circular definition, for it presupposes an understanding of cognitive processes, cognitive systems, and cognitive methodologies. The constructive way to look at this characterization, however, is to accept that the five notions—cognitive science, cognitive behaviour, cognitive processes, cognitive systems, and cognitive methodologies—are interdefinable. On this basis, then, how can we best proceed with the aim to specify cognitive processes (Fig. 1).

Let me outline the main line of argument of this paper: I am accepting a version of the multiple criteria account concerning cognitive processes, i.e. any description of cognitive processes needs to rely on a minimal understanding of cognitive science, cognitive behaviour, cognitive systems, and cognitive methodologies. But we have to give up the aim to work out a definition which could avoid every charge of circularity. Instead we should aim for a characterization of cognitive processes which

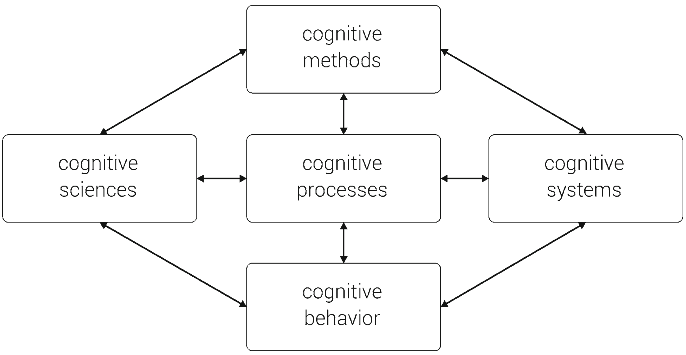


**Fig. 1** Interdependence of aspects of cognition

is sufficiently clear despite its interdependence on the other dimensions of cognition. Furthermore, the new characterization of cognitive processes should account for some central criteria of adequacy. To reach this aim, I proceed as follows: First, I account for and explicate the interdependence relation appealed to in characterizing cognitive processes (Sect. 2), and then I describe how a Wittgensteinian understanding of cognitive processes allows us to accept this interdependence and develop a fruitful characterization of cognitive processes (Sect. 3). One central aspect of the Wittgensteinian approach is the strategy of characterizing cognitive processes by typical examples (Sect. 4). This strategy is implemented by the core example of the comparator model (Sect. 5). Thus, I want to ground a sufficiently clear scientific understanding of cognitive processes on an example-based approach. The status of the proposal developed here is neither purely descriptive nor strictly normative but it is a contribution which combines conceptual considerations and empirical observations to anchor a scientific account.

## 2 Typical cognitive sciences, systems, behaviour and methods

If we are not in a position to define one of the related notions of cognition in a manner sufficiently independent from the others, must we then give up our hope to provide a characterization of cognitive processes? I don't think so. If each of the notions can be anchored by typical examples, we have a basis on which to develop the five notions via prototypical cases, and then each of these example-based notions can be used to sharpen our descriptions and theoretical understandings of what is going on in cognitive science. This requires that we give up our aim to develop an explicit definition of cognitive processes, and that we aim instead at a more modest example-based characterization influenced by our understanding of the neighbouring notions. Let us develop core examples for each of the four areas, before finally approaching the key area of cognitive processes.

There is general agreement that the typical cognitive sciences include psychology, philosophy, linguistics, anthropology, artificial intelligence, and neurosciences (Miller 2003, p. 143). The places of biology, medicine, and informatics are not clear and can only be clarified by considering other cognitive phenomena such as systems, behaviour, and methods (Fig. 2).

Is there a characteristic cognitive methodology? This question is closely connected to the problem of the identification of the typical disciplines in cognitive science, but it does not automatically follow that all the methods used in a typical cognitive science are characteristic of cognitive investigations. For example, researchers in philosophy make use of conceptual analysis or metaphysical speculation: both methods are typical for philosophy, but they are not typical cognitive methods. Typical cognitive methods are those which are especially useful for the interdisciplinary project of understanding the mind by discovering relevant empirical data or making progress in theory formation which explicitly accounts for the empirical basis. The discipline of philosophy also contributes typical methods to the cognitive sciences, such as the methods of systematic and integrative theory formation. Linguistics is concerned with the reconstruction of language-processing in humans through syntactic, semantic, and

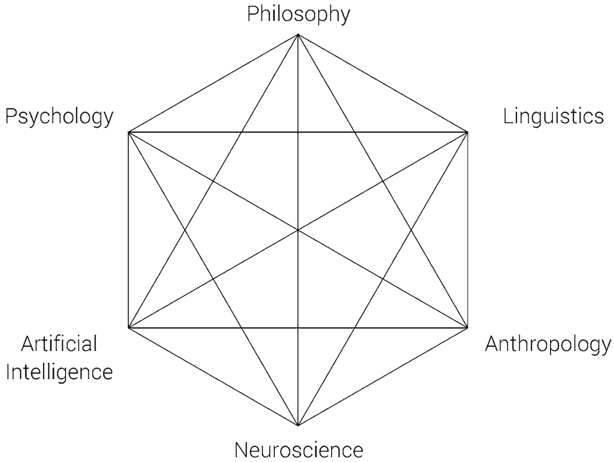


**Fig. 2** Typical disciplines of cognitive sciences (adopted from Miller 2003)

pragmatic analysis—a typical cognitive method; but it is also concerned with language documentation[1] aimed at preserving knowledge about languages which are threatened with extinction. This is obviously not a cognitive method, despite being typical for linguistics. Each typical discipline thus contributes some typical methods, but their typicality is dependent on the fact that they are relevant for making progress in understanding what we understand as cognitive processes. Thus, we return to the point that the five cognitive aspects are interdependent.

The *systems* typically targeted for cognitive investigations are humans, animals and robots. The status of plants with complex environment-dependent mechanisms is unclear, as is that of machines with complex internal programs. Again, to discuss these cases, we must consider other dimensions.

Typical cognitive *behaviour* is usually also called "intelligent behaviour": solving mathematical problems and memorizing events, but also practical tasks such as parking one's car at a specific spot, or using (at least minimally flexible) spatial navigation, e.g. of rats in a maze. Cognitive behaviour involves or is realized by cognitive processes. But not all behaviour of typical cognitive systems is cognitive behaviour—for example, we usually describe the knee-jerk reflex or the scratching of my itching arm as examples of non-cognitive behaviour in humans. Some complex movements like the opening and closing of a flower blossom are realized by chemical processes (or mechanisms) which do not deserve to be characterized as cognitive. Why not? Because these processes do not enable behaviour which is sufficiently complex; because such behaviour is not transferable to other typical cognitive systems, since we do not find it in animals or humans; and because technologically we can realize sensitivity to sunlight more easily through a coupled movement of a system than through building a cognitive system. Take, for example, the energy-saving houses called "Drehhaus",which are

[1] Language documentation is an important endeavour in times in which dialects and whole languages are rapidly dying out. An example is the Berkeley documentation: http://linguistics.berkeley.edu/research/field/index.

able to keep the glass front of the house facing in the direction of the sun.[2] Although the mechanism that realizes this movement involves quite complex hardware and software, it does not seem right to evaluate it as a cognitive mechanism, since a house is in no way a candidate for a cognitive system: its "behaviour" is extremely rigid, the underpinning design is based not on a typical cognitive science but on architecture, and a special cognitive method is not required to investigate houses. In all the interconnected dimensions it is far from typical.

## 3 A Wittgensteinian understanding of the concept of cognition

The answer to the question "What is a cognitive process?" depends on the account of "concepts" that one adopts (Kästner and Walter 2009). How do I situate the concept of COGNITIVE PROCESS within the spectrum of theories of concepts? The typicality considerations above presuppose that I am able to offer neither necessary nor sufficient conditions for being a cognitive process. Am I then forced to accept a prototype theory in a narrow sense? I do not think so. Although I accept that the only fruitful epistemic access that we have to characterize cognitive processes lies in offering typical examples, it remains underdetermined which ontology of concepts comes with this epistemic view. Concerning the remaining alternatives I want to remain neutral: it is possible that on the basis of a fruitful example-based characterization we will end up with nothing but a fruitful convention for using a concept given the actual state of science; or, indeed, we may in fact establish a natural kind term referring to cognitive processes as determined by nature. There are different ways of understanding natural kinds, for example as a kind determined by *one* key property [as with Putnam's (1975) account of water as determined by $H_2O$], or by a cluster of properties (Boyd 1999; Buckner 2015). The development of cognitive science will help us in the future to decide which understanding of the example-based strategy is best; but given the early stage of cognitive science it seems most reasonable to be modest, and to concentrate on convincing examples which at least *might* have the power to characterize an independent science for the time being. Here I adopt one of Wittgenstein's core proposals for thinking about concept when discussing the concept GAME[3]: the basis for describing the concept GAME is our use of the word "game" in the relevant contexts. This involves two steps, namely to characterize the interdependence of the concept GAME with other relevant concepts and then to describe typical examples while these typical examples are the basis for characteristic features which may result in the best theoretical characterization of the relevant concept. I proceed in an anal-

2 A description of this moving house can be found at: http://www.drehhaus.de/de/zweite-generation.php.

3 This illustrates Wittgenstein's thoughts about concept expressed by the word "game": "Consider for example the proceedings that we call 'games'. I mean board-games, card-games, ball-games, Olympic games, and so on. [...] Look at the parts played by skill and luck; and at the difference between skill in chess and skill in tennis. Think now of games like ring-aring-a-roses; here is the element of amusement, but how many other characteristic features have disappeared! And we can go through the many, many groups of games in the same way; can see how similarities crop up and disappear. And the result of this examination is: we see a complicated network of similarities overlapping and criss-crossing: sometimes overall similarities, sometimes similarities of detail" (Wittgenstein 1967, PU § 66).

ogous fashion for the case of the concept COGNITIVE PROCESS, characterizing concepts by describing their use, first highlighting their interdependence with relevant neighbouring concepts (as I have done in the previous section), and second describing typical examples (which will be done in Sect. 5). This enables us to develop the framework for the best scientific characterization of cognitive processes which is the aim of this paper. The example-based approach establishes a fruitful notion of cognitive processes, but leaves the ontology open for future decisions on the basis of further evidence and integrative theory formation.

## 4 An example-based approach to cognitive processes in the light of some criteria of adequacy

What are typical examples of cognitive processes? After introducing the idea of interdependence (see Fig. 1) we can offer a first rough characterization, not to be understood as providing a definition, but rather as constraining our search for typical examples. *Cognitive processes* are processes of information transfer that typically take place to connect multiple (or complex) informational inputs to form a minimally flexible cognitive system with a spectrum of minimally flexible behavioral outputs, where these processes typically involve at least a minimal level of one of the following paradigmatic processes as described in some cognitive sciences with a cognitive method: perception, memory, learning, emotion, intentionality, self-representation, rationality, and decision-making or something relevantly similar to it. This description has two features. It accounts for the interdependence of aspects of cognition, and it suggests that we go "minimal" in response to the demands of each dimension.

Why should we go minimal in describing typical examples? This is a first step in developing three criteria of adequacy: cognitive processes should cover more than conscious processes and more than those characterized by folk-psychological explanations. Let me illustrate this: the notion of cognitive processes can be made fruitful by using it to pick out something that differs from *conscious* processes. Here I appeal to a background intuition that we use these notions differently, and that we have at least some agreement on our uses of the terms "conscious phenomena" and "cognitive phenomena"; namely, a phenomenon is conscious for a cognitive system if in typical cases it is connected with some experience which has a phenomenal character that we understand as a subjective and private experience of this system. We know that humans can have conscious emotions, perceptions etc., but they can also undergo unconscious perceptions, e.g. in cases of visual agnosia (Milner and Goodale 1995). Such a phenomenon is not conscious, yet it is still cognitive. We can see this by considering cases of a person suffering from visual agnosia who still processes typical information (such as the direction of a letter box, and despite not being consciously aware of it) where this information still plays the standard role in folk-psychological explanations—"Why is D.F. able to post the letter into the box?" "Because D.F. has a perceptual state with the content that the letter box now lies in a specific direction". Again, this representational state deserves to be called a cognitive state because it plays its standard role, just like perceptual information. We can use the same example to show that cognitive processes should involve more than just the processes described by folk-psychological explanations. In a more elaborate folk-psychological explanation of her action we might say that D.F. successfully posted the letter into the box because she wanted to post the letter and she had (unconsciously) seen the direction of the letter box. But in this case we know even more about the relevant neural mechanism, such that the folk-psychological explanation can be substituted by it: the relevant process involves a perception-based activation of the dorsal pathway, which explains the behaviour of D.F. In many cases of mechanisms of attention, memory, or learning, we describe the relevant psychological effect but cannot offer a folk-psychological explanation, e.g. some hippocampal processes produce the storage of episodic memory (Eichenbaum 2013). The divide between the conscious and unconscious phenomena which guide behaviour leaves room for phenomena to be called "cognitive" if the processes are part of a mechanism that enables us to explain behaviour, even though the mechanisms are neither typically conscious nor referred to as part in a folk-psychological description. A second criterion of adequacy demands that the characterization of cognitive processes should be explanatory fruitful. I am presupposing a background view comprising more and more specific processes, starting with physical processes and ending up with complex conscious processes. In such a framework the presupposition of cognitive processes is explanatorily fruitful if it offers a special *epistemic* role for cognitive processes which are less specific than conscious processes, but more specific than biochemical or physical processes. Adopting a functionalist perspective, we should distinguish the biochemical and physical processes from the cognitive processes if we observe multiple realizations of the same cognitive processes on different materials. The third criterion of adequacy demands that we should characterize cognitive processes such that they can be used to explain the behaviour of humans as well as animals and robots (or other sufficiently complex AI systems, see Fig. 3).

To summarize the theoretical stance of my account: I accept the multiple criteria account for characterizing cognitive processes including the observation of the systematic interdependence of concepts expressed by using the adjective "cognitive …". Given the latter, it is only possible to come to a scientific characterization of cognitive processes if we start off with a rough idea constraining the relevant neighbouring

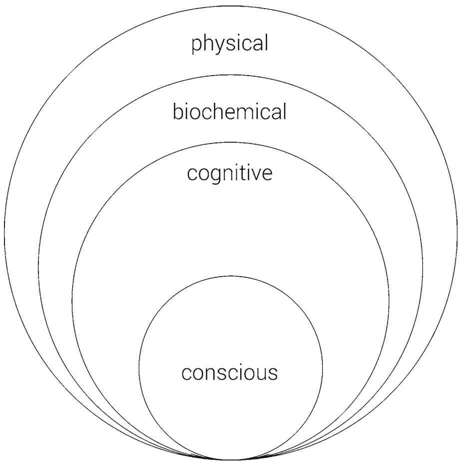


**Fig. 3** Situating cognitive processes for a fruitful explanatory perspective

concepts of cognitive systems, cognitive sciences, cognitive behaviour, and cognitive methods. Here I suggest that we start with a minimal understanding of these concepts on the basis of intuitions guided by actual practices. On the basis of such a framework, we need to find convincing examples of cognitive processes which support the working hypothesis and show that cognitive processes demark a useful scientific category. The hypothesis may be worth being repeated here: *Cognitive processes* are processes of information transfer that typically take place to connect multiple (or complex) informational inputs to form a minimally flexible cognitive system with a spectrum of minimally flexible behavioral outputs, where these processes typically involve (at least a minimal level of) one of the following paradigmatic processes as described in certain cognitive sciences with a cognitive method: perception, memory, learning, emotion, intentionality, self-representation, rationality, and decision-making or something relevantly similar to it. To make this more concrete: we need to show on the basis of examples that a notion of cognitive processes enables us to explain minimally flexible behaviour with a core mechanism sometimes shared by animals, robots, and humans, where this mechanism is also involved in typical examples of cognitive sciences like perception, memory etc.[4] Given this perspective it is explanatory fruitful to distinguish cognitive processes from conscious processes as well as physical and biological processes. Otherwise we do not need such a category in science at all. Here I present such an example in detail, namely the *comparator model* (von Holst and Mittelstaedt 1950; Blakemore et al. 1998) which is a mechanism (i) below the level of consciousness and folk-psychological explanation, (ii) that enables us to explain a variety of phenomena, and (iii) that is used in humans, animals, and AI systems. In addition to the core example, the comparator model, I also make brief mention of certain others in order to defend the claim that we need to presuppose cognitive processes as a scientific category; the example-based approach is then developed in order to show that the examples do give us the best starting point for a future search for a theory of cognitive processes. This is a modest aim, but given the early stage of cognitive science it seems to be the best we can do now.

## 5 The core example: the comparator model

### 5.1 Explaining goal-directed movements, tickling experiences, and the feeling of agency

Is there evidence for *cognitive processes* which are neither typically conscious or folk-psychological nor which typically pertain only to processes that can be described as purely physical, chemical, or biological? Before presenting positive examples, let us look at a borderline case of a mechanism underlying rigid behaviour which we may want to exclude, namely the mechanism underlying the ability of a bacterium to swim across a gradient. Macnab and Koshland (1972) demonstrated that a bacterium

[4] This implies that if we have a mechanism which only explains very rigid behavior without playing any role in producing more complex flexible behaviour, then it is not fruitful to count it as a cognitive mechanism.

(*Salmonella* strain) does not, as one might have guessed, register a spatial difference in nutrient levels across its "body", but registers a temporal difference in the nutrient levels and maintains its current direction of movement as long as an increase in nutrient level over time is detected. While this is quite a sophisticated mechanism, which might even be described as involving a comparison mechanism at the cellular level, it is questionable whether this is a typical cognitive mechanism. It is a mechanism which triggers rigid behavior, seems only to be used for this purpose, and is not integrated into any further abilities. Thus, it is better to classify it as a pure biological mechanism. The first positive candidates are basic learning processes like association or conditioning processes, for we know that what happens in such cases is often implicit learning, at least in cases like Pavlov's dog. Such learning is unconscious, but involves much more than just the pairing of two stimuli. Rescorla (1988) shows that conditioning can be adequately described as "the learning of relations among events so as to allow the organism to represent its environment" (Rescorla 1988). And this way of accounting for conditioning allows us to understand its role in central psychological processes like emotion and motivation (Rescorla 1988). The basic unconscious conditioning process is not just a basic and rather rigid mechanism but is also integrated into central psychological processes, and can be observed in humans and animals as well as implemented in robots: thus it is a good candidate for a cognitive process. Another example is the mirror neuron mechanism. We know from the work of Gallese and Rizzolatti (Gallese et al. 1996; Rizzolatti and Craighero 2004) that the same small group of neurons fire when I observe a basic action like grasping a glass as when I actually do it myself. Mirror neurons fire in both cases involving basic goal-directed movement, i.e. in an observation mode and in an execution mode. This was an interesting discovery of a mechanism relevant for basic processes in goal-directed grasping. It was discovered in monkeys, but then also shown to have an equivalent in humans when we undertake a goal-directed action and observe it in others (Keysers and Gazzola 2010), and is clearly also involved when undergoing or recognizing disgust (Wicker et al. 2003) and pain (Hutchison et al. 1999). These undisputed observations lead to the speculation that mirror neurons are the central neural correlate of social cognition in general (Gallese and Goldman 1998).[5] The latter claim has been subject to intense criticism (Jacob 2008; Heyes 2010; Spaulding 2013; Cook et al. 2014). It remains an open question to which extend mirror neurons are relevant for social cognition; yet there is sufficient consensus that they form the neural correlate of emotional empathy (Jabbi et al. 2007) (while it is unclear whether they are relevant for cognitive empathy). Thus, mirror neurons are also relevant for central psychological processes like emotions.

Are there even more basic mechanisms which are on the one hand part of basic evolutionary processes in simple animals and on the other hand part of central psychological processes in humans? If there are, this would support the claim that cognitive processes are a fruitful scientific category. We thus stand in need of an even more convincing example, which has to be spelled out in detail with respect to all the dimensions

[5] This claim is defended together with the background theory that social cognition is mainly mental simulation as described in simulation theory (Goldman 2006); but it is criticized by alternative accounts such as theory-theory (Baron-Cohen 1995; Gopnik 1993), interaction theory (Gallagher 2001; Gallagher and Hutto 2008), and person model theory (Newen and Schlicht 2009; Newen 2015).

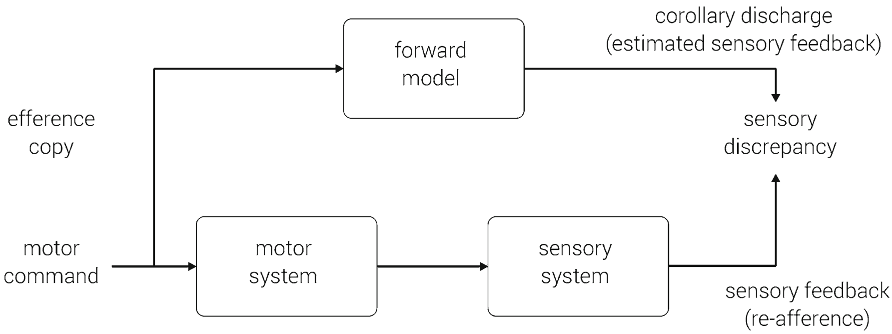


**Fig. 4** The basic comparator mechanism (adapted from von Holst and Mittelstaedt 1950)

of cognition introduced above: and in this regard I suggest that the *comparator* (or feedforward) mechanism can do the job (Fig. 4).

In its classical and simplified version, the comparator mechanism (von Holst and Mittelstaedt 1950; Blakemore et al. 1998) involves a motor command being initiated on the basis of an environmental input which triggers two routes of further processes: one is the classical route of triggering a movement by activating the motor system which is connected with sensory feedback from the actual moving body; and a second route triggers the *expected* state (given a successful motor command)—thus we can also speak of the estimated sensory feedback. The essence of the process is the comparison of the actual sensory feedback using the estimated or expected feedback, which may result in either a match or a mismatch. Let us illustrate the comparator model with a simple goal-directed movement. Perhaps I aim to grasp an apple. My motor system triggers a movement of my right arm which comes with actual sensory feedback signalling the actual position of the arm; and it also triggers an expected feedback (sometimes also called the efference copy). If there is a match the movement is completed. If there is a mismatch, perhaps because the arm is passively moved by another person in the wrong direction, then I automatically adjust the motor command to complete the goal-directed movement which should result in grasping the apple. Online correction of our goal-directed movements is one important application of the comparator model mechanism. Why is this a typical cognitive process?

It is a typical cognitive process because it is typical in all the dimensions described in our account of the interdependencies of cognitive features. Before discussing each dimension, let me briefly note that the process is clearly a candidate for a cognitive process which is neither typically conscious nor typically folk-psychological, and thus it satisfies the first criterion of adequacy: as humans we are not conscious of online corrections of our movements. Nor is this process part of folk-psychological considerations. Now, though, I need to argue that it is complex enough to count as cognitive. Figure 4 shows the basic comparator mechanism as it allows online correction.

In the background of online corrections of movement we normally presuppose (in the case of humans) an intention to act. The basic mechanism presupposes neither a conscious intention nor an intention in the sense of folk psychology, but only a motivational trigger for a goal-directed action such as an urge to grasp food. Are there applications of the mechanism which show that it is explanatorily complex enough to not be classified as a pure biochemical process? The same mechanism offers an elegant explanation of the phenomenon that we cannot tickle ourselves (Blakemore, Wolpert, and Frith 2000). If we want to tickle ourselves, we aim at touching a specific part of our body, e.g. a spot on the left arm. Given the comparator model, we develop an efference-copy of the estimated sensory feedback expecting to be touched at the very part of the left arm. If we touch ourselves successfully, this produces a perfect match of actual and estimated feedback. However, as the feeling of tickling essentially involves some kind of *unpredicted* touching, this is clearly not possible through normal self-touching, which is always predicted—we can only be tickled by someone else. One exception is possible according to this model, namely if we could arrange some type of self-touch such that the touch of my hand is not predicted by some efference copy, i.e. by the estimated sensory feedback. This can actually be done: if we introduce a metal construction such that a lever is moved by the right hand and the construction unfolds behind my head such that my left hand is touched at an unexpected spot by a metal tip at the end of the construction which I move with my right hand, then I can actually tickle myself (Blakemore et al. 1999). This is consistent with and predicted by the comparator model: only if the actual sensory feedback is unexpected can it produce a tickling sensation, and normal self-tickling cannot produce this whereas the special mechanism can. Thus, in addition to online correction of movement, we have found another application, i.e. the case of self-tickling.

There are also applications of the comparator mechanism in the direction of more high-level phenomena, such as the mental state of *having a feeling of agency*. We experience a normal goal-directed action like grasping a glass to take a drink as being combined with a feeling of agency—that is, I experience that it is *me* doing it, not someone else. According to Frith et al. (2000) this can also be explained by the comparator model. To do this, the model needs an explicit extension such that we start with an intention to act. Perhaps I want to pick up my mobile phone which is lying on the table. The dual-triggering process by motor command begins, and culminates in a comparison between actual and expected feedback. In the case of a match I end up with a feeling of agency. In the case of a passive movement in which someone takes my hand, takes the mobile phone, and puts it into the hand, I may end up with the same state of having the mobile phone in my hand yet I do not have a sense of agency combined with the movements that my hand was undergoing. It is a matter for ongoing research whether the comparator model is sufficient to account for the complex phenomenon of agency. In papers written in cooperation with other researchers I have argued that we have to distinguish between the feeling of agency and the judgment of agency (Synofzik et al. 2008a). The comparator model is the key to explain the feeling of agency (although it involves some additional features), while in the case of an explicit judgment of agency many more factors become relevant to producing the phenomenon, such as situational context, additional semantic information, background thoughts etc. (Synofzik et al. 2008a, b). If we accept that the comparator mechanism is a key mechanism for generating a feeling of agency in the case of goal-directed movements, we have intuitively reached the level of typical psychological phenomena which may even be combined with conscious experience in some cases. Nevertheless, the core mechanism is a rather basic cognitive

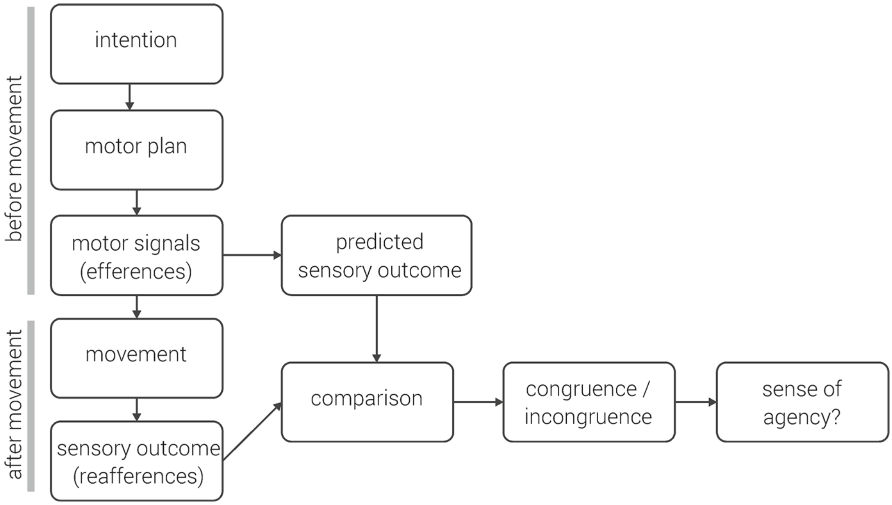


**Fig. 5** The adapted comparator mechanism to explain a sense of agency

mechanism which is not specific to or typical for producing conscious phenomena (Fig. 5).

### 5.2 Arguing for the typicality of the comparator mechanism as a cognitive process

After this characterization of a first group of three applications which manifest the needed explanatory fruitfulness (criterion 2), I now discuss the four interconnected aspects of cognitive processes (Fig. 1). First, we have seen that the comparator mechanism satisfies the demand that it account for *cognitive behaviour*, i.e. *minimal flexible behaviour*, for it accounts inter alia for online correction of goal-directed movement, the possibility to tickle others and the impossibility to tickle oneself and even more. Second, since we have already illustrated three applications in the realm of human behaviour, the mechanism is part of the group of psychological mechanisms. Thus we have a *typical cognitive science* which makes heavy use of this mechanism, and thus it is part of a standard psychological explanation strategy. It follows, third, that explanations with the comparator model are thereby psychological methods of explanation. Moreover their typical use in explaining not only folk-psychological or conscious phenomena but also a wider range of phenomena indicates that it would be best to describe this explanatory strategy as *a typical cognitive methodology.* The fourth and final question, then, is whether the comparator mechanism is or can be used to explain, predict, or shape the behaviour of different typical *cognitive systems*. So far we have discussed only applications to human behaviour, and we still need to discuss animals and robots. This would enable us to satisfy the third criterion of adequacy.

Let us dwell on the case of animals, since this allows us to observe even more key applications. Here I can report that—interestingly enough—the mechanism was first discovered through explaining the optokinetic effect in flies (von Holst and Mittelstaedt 1950). How can a fly distinguish whether it or the environment is moving, since the resulting activation on the retina is the same? The comparator model offers an answer: a cognitive system only produces an efference copy if it moves itself. Only in this case can the result be a match between expected and actual sensory feedback, enabling the fly to land on an object, or, in the case of mismatch, to correct its movements. In the case of movement of the environment the efference copy is lacking. Thus there is a clear difference in information processing in the cases of self-movement and environment-movement, despite the indistinguishable activation of the retina. The same mechanism is developed to distinguish visually between figure and background in bees: given a current state of sensory input, a bee instantiates a motor command that makes it move in a certain direction. On this basis it can register a match between estimated and actual sensory feedback; thereby the direction of movement is registered and adjusted. If there is a mismatch between expected and actual feedback, the figure–ground model is updated by changing the motor command with respect to the actual feedback (Kern et al. 1997).

One consequence of this comparator model is that it engenders a new way of thinking about the interaction between perception and action: motor activity influences the actual sensory feedback (called also reafference, see von Holst and Mittelstaedt 1950). Thus action directly influences sensation and vice versa, since the actual feedback, a sensation, modifies or corrects the motor command if necessary. This can be regarded as a form of active sensing strategy, which is also characterized in recent discussions as the perception–action cycle. A further application of the comparator model understood in this way can be seen in the active sensing of objects by weak electric fish (Hofmann et al. 2013): weak electric fish have a kind of a sixth sense whereby they register objects by actively producing weak electric fields: elephantnose fish can be trained to recognize different properties of objects based on electrolocation. These properties include the size, distance, impedance, and shape of an object based solely on electric image properties. It is important for these fishes to distinguish their own produced electric fields from those of other electric fishes. And this can be easily accounted for with the comparator model, where the efference copy provides the key to the difference. The same principle is used in echolocation in bats, who need to distinguish their emitted waves from those of other bats to orient themselves in the environment (Hofmann et al. 2013). The same mechanism is also analyzed in detail in singing crickets, where Poulet and Hedwig (2006) have identified the cellular basis of a basic comparator process that is indispensable in order to distinguish self-generated sensory feedback from external information. Thus there is a great deal of evidence from cases of simple vision-based movements, as well as from spatial navigation and species-member registration, both relying on auditory signals. And if we now turn to apes, who are able to do the same type of basic goal-directed actions as humans, and given the similarity of their brain organisation to ours, it is usually suggested that their goal-directed grasping behaviour is shaped by the same mechanism as in humans, including the comparator mechanism.

What about applications in artificial intelligence and robotics? In the context of AI, comparator models are often called "feedforward" models. Feedforward models have been standard in AI at least since the 1990s, in modelling robot arms (e.g. Wells et al. 1990) and in controlling industrial robots (Grotjahn 2002). In recent developments they have also been a constitutive element of starting to model social understanding in artificial systems (DelRose et al. 2011). This should be sufficient to conclude that the comparator model has important applications in AI as well. Thus we find the comparator mechanism in humans, animals, and robots: which completes our consideration of typicality. The comparator mechanism really is a typical cognitive process. There have also been discussions of its contribution to other high-level mental phenomena, e.g. self-recognition (Tsakiris et al. 2005), but as I described for the phenomenon of judgment of agency (see above), the comparator mechanism is losing its central role in those cases since a lot of other important features are involved too. This is also the case in other high-level psychological phenomena which I will illustrate by discussing the phenomenon of the authorship of thoughts. This will complete the picture of the role of paradigmatic basic cognitive processes: they are typically involved in implementing basic psychological phenomena, they can play a central role in several other (mid-level) psychological phenomena. Furthermore, an adequate embedding of these cognitive mechanisms can lead to the realization of high-level psychological phenomena. But the latter are strongly multi-factorial, and the core cognitive mechanism is only one of many relevant components in their complex realization patterns.

### 5.3 Further applications in humans: the attribution of authorship of thoughts

There is a tendency in parts of the literature to claim that the comparator mechanism is also able to account for the authorship of thoughts, i.e. to explain why a thought which I experience now is my thought and not a thought inserted by someone else into my head. Thought ascription is a typical mental phenomenon, and thus this case provides a challenge. The breakdown of normal attribution of the authorship of thoughts happens in the case of thought insertion. This is a well-known and typical syndrome of schizophrenia. Can this be fully explained by a version of the comparator model? First, we should note that even according to researchers who aim to develop an explanation on this basis, a simple comparator model is not enough; it would at least be necessary to appeal to iterative uses of different comparators (Frith et al. 2000). Furthermore, there is a main line of interpretation as regards these phenomena which maintains that we need to distinguish two factors in explaining the symptoms: first, an experiential factor, i.e. a disturbance at the level of the sensory system; and second, and in addition, a (local) breakdown of the ability to attribute rational attitudes on the basis of the sensory input (Davies et al. 2001). If this view is correct, we need many more factors to explain thought insertion than merely an advanced comparator mechanism (Vosgerau and Newen 2007). Last but not least, there is the philosophical claim that all everyday thoughts (understood as events of thinking) are a product of a comparator mechanism which treats thoughts in fashion that is a strictly parallel to motor processes (Campbell 1999). As a colleague and I have extensively argued, this view is not tenable (Vosgerau and Newen 2007). Thus, we have good reasons not to overextend the role of the comparator mechanism. It may be an important contributing component of thought insertion on the level of disturbed sensory processing, but the two-factor theory indicates that it is essentially dependent on several other factors, including, at least, the breakdown of the capacity for rational attribution of attitudes.

## 6 Theoretical evaluation and conclusion

In so far as the state of the cognitive sciences allows, having started with some standard candidates, I hope I have presented a convincing example of a cognitive process: the comparator mechanism. It satisfies the interdependency dimensions, and is typical in the sense of accounting for cognitive phenomena in general and not only conscious or folk-psychological phenomena; and it is evolutionary old, being implemented in flies. Can we go further, and establish clear necessary and sufficient conditions on the basis of which to define cognitive processes? I do not think so. The interdependence observation implies that the notion of cognitive processes is dependent on many aspects, including actual practice in the typical cognitive sciences, the development of specific research methods, etc., and these all undergo the normal processes of change characteristic of the history of any science. Given the immature state of cognitive science as an interdisciplinary endeavour, we should not expect to be able to characterize more than a typical cluster of processes, even though a good way to do this is still to offer convincing paradigm examples of cognitive processes as a basis for typicality comparisons. Can we go so far as to make ontological claims, at least for the paradigm examples? Are we justified in claiming that cognitive processes constitute a natural kind, as suggested by e.g. Buckner (2015)? All we now know is clearly compatible with the view that we are dealing mainly with biological processes, which may constitute natural kinds even while their functional roles vary when used in organisms which need to meet rather different environmental challenges, especially over evolutionary timescales. Well-known observations concerning the plasticity of the brain (Hübener and Bonhoeffer 2014) show how wide this degree of variation can be. Thus, the same biological mechanism can be used to realize rather different cognitive functions, which certainly is a challenge for the natural kind view of cognitive processes. I therefore leave it open for future developments within the interdisciplinary working community to determine whether the example-based characterization ends up only identifying a class of mechanisms clustered by conventional practices—a class which happens to be particularly fruitful at the beginning of the twentyfirst century—or whether the examples do in fact identify a natural kind or a property cluster. We will have to await progress both in the empirical data and in theory formation to take a principled stance.

**Acknowledgments** For helpful comments, I would like to thank Francesco Marchi, Pascale Willemsen, and two anonymous referees.

## References

Adams, F., & Aizawa, K. (2008). *The bounds of cognition*. Oxford: Blackwell.

Baron-Cohen, S. (1995). *Mindblindness. An essay on autism and theory of mind*. Cambridge, MA: MIT Press.

Blakemore, S. J., Goodbody, S. J., & Wolpert, D. M. (1998). Predicting the consequences of our own actions: The role of sensorimotor context estimation. *Journal of Neuroscience*, *18*(18), 7511–7518.

Blakemore, S. J., Frith, C., & Wolpert, D. M. (1999). Spatio-temporal prediction modulates the perception of self-produced stimuli. *Journal of Cognitive Neuroscience*, *11*(5), 551–559.

Blakemore, S. J., Smith, J., Steel, R., Johnstone, C. E., & Frith, C. D. (2000). The perception of self-produced sensory stimuli in patients with auditory hallucinations and passivity experiences: Evidence for a breakdown in self-monitoring. *Psychological Medicine*, *30*(5), 1131–1139.

Boyd, R. (1999). Kinds, complexity and multiple realization. *Philosophical Studies*, *95*(1), 67–98.

Brooks, R. A. (1991). Intelligence without representation. *Artificial Intelligence*, *47*, 139–159.

Buckner, C. (2015). A property cluster theory of cognition. *Philosophical Psychology*, *28*, 307.

Campbell, J. (1999). Schizophrenia, the space of reasons, and thinking as motor process. *The Monist*, *84*, 609–625.

Clark, A., & Chalmers, D. (1998). The extended mind. *Analysis*, *58*, 10–23.

Cook, R., Bird, G., Catmur, C., Press, C., & Heyes, C. (2014). Mirror neurons: From origin to function. *Behavioral and Brain Sciences*, *37*, 177–241. doi:10.1017/S0140525X13000903.

Davies, M., Coltheart, M., Langdon, R., & Breen, N. (2001). Monothematic delusions: Towards a two-factor account. *Philosophy, Psychiatry, and Psychology*, *8*, 133–158.

DelRose, M., Wagner, C., & Frederick, P. (2011). Evidence feed forward hidden Markov model: A new type of hidden Markov model. *International Journal of Artificial Intelligence and Applications*, *2*(1), 1–19. doi:10.5121/ijaia.2011.2101.

Eichenbaum, H. (2013). Memory on time. *Trends in Cognitive Sciences*, *17*(2), 8–81. doi:10.1016/j.tics.2012.12.007. Epub 2013 Jan 12.

Frith, C. D., Blakemore, S., & Wolpert, D. M. (2000). Explaining the symptoms of schizophrenia: Abnormalities in the awareness of action. *Brain Research Review*, *31*(2–3), 357–363.

Gallagher, S. (2001). The practice of mind: Theory, simulation, or interaction? *Journal of Consciousness Studies*, *8*(5–7), 83–107.

Gallagher, S. (2005). *How the body shapes the mind*. New York: Oxford University Press.

Gallagher, S., & Hutto, D. D. (2008). Understanding others through primary interaction and narrative practice. In J. Zlatev, T. Racine, C. Sinha, & E. Itkonen (Eds.), *The shared mind: Perspectives on intersubjectivity* (pp. 17–38). Amsterdam: John Benjamins.

Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain*, *119*, 593–609.

Gallese, V., & Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends in Cognitive Science*, *1/2*(12), 493–501.

Goldman, A. I. (2006). Simulating minds: The philosophy, psychology, and neuroscience of mindreading. New York: Oxford University Press.

Gopnik, A. (1993). How we know our minds: The illusion of first-person knowledge of intentionality. *Behavioral and Brain Sciences*, *16*(1), 1–15, 90–101. doi:10.1017/S0140525X00028636.

Grotjahn, M. (2002). Model-based feedforward control in industrial robotics. *The International Journal of Robotics Research*, *21*, 45–60.

Heyes, C. (2010). Where do mirror neurons come from? *Neuroscience and Biobehavioral Reviews*, *34*, 575–583.

Hofmann, V., Sanguinetti-Scheck, J. I., Künzel, S., Geurten, B., Gómez-Sena, L., & Engelman, J. (2013). Locomotion and sensing of weakly electric fish: Sensory flow shaped by active sensing: Sensorimotor strategies in electric fish. *Journal of Experimental Biology*, *216*, 2487–2500. doi:10.1242/jeb.082420.

Hübener, M., & Bonhoeffer, T. (2014). Neuronal plasticity: Beyond the critical period. *Cell*, *159*(4), 727–737. doi:10.1016/j.cell.2014.10.035.

Hurley, S. (1998). *Consciousness in action*. Cambridge, MA: Harvard University Press.

Hutchison, W., Davis, K., Lozano, A., Tasker, R., & Dostrovxky, J. (1999). Pain related neurons in the human cingulate cortex. *Nature Neuroscience*, *2*, 403–405.

Jacob, P. (2008). What do mirror neurons contribute to human social cognition? *Mind & Language*, *23*(2), 190–223.

Jabbi, M., Swart, M., & Keysers, C. (2007). Empathy for positive and negative emotions in the gustatory cortex. *NeuroImage*, *34*(4), 1744–1753.

Kästner, L., & Walter, S. (2009). What is cognition?—functionalism and extended cognition. In V. A. Munz, K. Puhl, & J. Wang (Eds.), *Language and world. Preproceedings of the 32nd International Wittgenstein Symposium.* Kirchberg am Wechsel, Österreich: Ontos.

Kern, R., Egelhaaf, M., & Srinivasan, M. V. (1997). Edge detection by landing honeybees: Behavioural analysis and model simulations of the underlying mechanism. *Vision Research, 37*(15), 2103–2117.

Keysers, C., & Gazzola, V. (2010). Social neuroscience: Mirror neurons recorded in humans. *Current Biology*, *20*(8), R353–R354.

Macnab, R. M., & Koshland, D. E. (1972). The gradient-sensing mechanism in bacterial chemotaxis. *PNAS*, *69*, 2509–2512.

Menary, R. (2010). *The extended mind*. Cambridge: M.I.T. Press.

Miller, G. A. (2003). The cognitive revolution: A historical perspective. *Trends in Cognitive Sciences*, *7*(3), 141–144.

Milner, A. D., & Goodale, M. A. (1995). *The visual brain in action*. New York: Oxford University Press.

Newen, A. (2015). Understanding others: The person model theory. In T. Metzinger, & J. M. Windt (Eds). *Open MIND.* Frankfurt a.M.: MIND Group, article 26, 1–28.

Newen, A., & Schlicht, T. (2009). Understanding other minds. A criticism of Goldman's simulation theory and an outline of the person model theory. *Grazer Philosophische Studien*, *79*(1), 209–242.

Noe, A. (2004). *Action in perception*. Cambridge, MA: MIT Press.

Poulet, J. F. A., & Hedwig, B. (2006). The cellular basis of a corollary discharge. *Science*, *311*, 518–522.

Putnam, H. (1975). The meaning of 'meaning'. In H. Putnam (Ed.), *Mind, language and reality. Philosophical Papers* (Vol. 2, pp. 215–271). Cambridge: Cambridge University Press.

Quine, W. V. O. (1951). Two dogmas of empiricism. *Philosophical Review*, *60*(1), 20–43.

Rescorla, R. A. (1988). Pavlovian conditioning. It's not what you think it is. *American Psychologist*, *43*(3), 151–160.

Rizzolatti, G., & Craighero, L. (2004). The mirror neuron system. *Annual Review of Neuroscience*, *27*, 169–192.

Robbins, P., & Aydede, M. (2009). *The Cambridge Handbook of Situated Cognition*. Cambridge: Cambridge University Press.

Spaulding, S. (2013). Mirror neurons and social cognition. *Mind and Language*, *28*(2), 233–257.

Synofzik, M., Vosgerau, G., & Newen, A. (2008a). Beyond the comparator model: A multifactorial two-step account of agency. *Consciousness & Cognition*, *17*, 219–239.

Synofzik, M., Vosgerau, G., & Newen, A. (2008b). I move, therefore I am: A new theoretical framework to investigate agency and ownership. *Consciousness & Cognition*, *17*, 411–424.

Thagard, P. (2010). Cognitive science. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2012 Edition). http://plato.stanford.edu/archives/fall2012/entries/cognitive-science.

Tsakiris, M., Haggard, P., Franck, N., Mainy, N., & Sirigu, A. (2005). A specific role for efferent information in self-recognition. *Cognition*, *96*, 215–231.

von Holst, E., & Mittelstaedt, H. (1950). Das Reafferenzprinzip. *Naturwissenschaften*, *37*, 464–476.

Vosgerau, G., & Newen, A. (2007). Thoughts, motor actions, and the self. *Mind and Language*, *22*, 22–43.

Wells, R., Schueller, J. K., & Tlusty, J. (1990). Feedforward and feedback control of a flexible robotic arm. *IEEE Control Systems Magazine*, *10*, 9–15.

Wicker, B., Keysers, C., Plailly, J., Royet, J. P., Gallese, V., & Rizzolatti, G. (2003). Both of us disgusted in my insula: The common neural basis of seeing and feeling disgust. *Neuron*, *40*(3), 655–664.

Wittgenstein, L. (1967). *Philosophical Investigations* (G. E. M. Anscombe, Trans.). Oxford: Blackwell (original in German from 1953).