Journal of Artificial Intelligence and Consciousness
Vol. 8, No. 1 (2021) 1–14
© World Scientific Publishing Company
DOI: 10.1142/S2705078521500077

Feeling Functional: A Formal Account of Artificial Phenomenology

Tarek R. Besold*

Neurocat GmbH, Rudower Chaussee 29
12489 Berlin, Germany
tb@neurocat.ai

Lorijn Zaadnoordijk

Trinity College Institute of Neuroscience
Trinity College Dublin,
Dublin 2, Ireland
l.zaadnoordijk@tcd.ie

David Vernon

Institute for Artificial Intelligence, University Bremen,
Am Fallturm 1, 28359 Bremen, Germany
david@vernon.eu

Received 1 November 2020
Accepted 13 December 2020
Published 16 February 2021

For humans, phenomenal experiences take up a central role in their daily interaction with the world. In this paper, we argue in favor of shifting phenomenal experiences into the focus of cognitive systems research and development. Instead of aiming to make artificial systems feel in the same way humans do, we focus on the possibilities of engineering capacities that are functionally equivalent to phenomenal experiences. These capacities can provide a different quality of input, enabling a cognitive system to self-evaluate its state in the world more effectively and with more generality than current methods allow. We ground our general argument using the example of the sense of agency. At the same time, we reflect on the broader possibilities and benefits for artificial counterparts to human phenomenal experiences and provide suggestions regarding the implementation of functionally equivalent mechanisms.

Keywords: Cognitive systems; phenomenology; embodiment; sense of agency.

1. Introduction

For humans, phenomenal experiences are a defining element of many interactions with the surrounding world. The salient part of turning one's face towards the sun is

[*] Corresponding author.

not the abstract registration of the sunbeams but the pleasant quality of feeling the sun on one's skin. The salient part of putting one's hand on the hot stove is not the abstract registration of a pain signal but the excruciating quality of feeling the latter causes, which usually leads to an immediate withdrawal of the hand. The salient part of holding one's child is not the abstract tactile registration and conceptual realization but the affectionate quality of feeling that comes with it. Phenomenal experiences provide a different quality of input to cognition as compared to non-phenomenal perception (i.e., abstract registration of stimuli from the environment). While the presence of phenomenal qualities in our everyday cognition does not always receive our active attention, their central role becomes clear when trying to imagine their absence. Cognitive capacities that rely in parts on phenomenal experiences include, for instance, learning [Mandler, 1989], social interaction [Gallagher, 2004; Gallese *et al.*, 2007], prospection [Gilbert and Wilson, 2007], and ethical behaviors [Keltner *et al.*, 2006; Torrance, 2008].

Phenomenal experiences are conceptually closely tied to notions such as consciousness and the self. Phenomenology has therefore been a popular topic of theoretical and empirical investigation across different disciplines including philosophy [Bayne and Montague, 2011; Chalmers *et al.*, 2004; Crane, 2003; Gallagher and Zahavi, 2013; Pacherie, 2008], as well as cognitive science and neuroscience [Varela *et al.*, 2016; Dehaene and Naccache, 2001; Haggard and Clark, 2003; Lamme, 2006] but — bar a few notable exceptions such as Chella and Manzotti [2011], Sloman and Chrisley [2003], Froese and Ziemke [2009] and Vernon and Furlong [2007] — has widely been ignored in the field of artificial intelligence (AI). We argue in favor of changing this, suggesting to shift phenomenology into the focus of cognitive systems research and development. Among others, phenomenology can, in a similar fashion to its function in biological systems, facilitate the self-evaluation of a systems state in the world. This in turn aids learning about and interacting with the physical world and other agents. Furthermore, adopting a phenomenological stance introduces a different set of assumptions about the nature of cognition and intelligence, and provides a foundation for enactive AI [Froese and Ziemke, 2009] and enactive embodied cognition [Varela *et al.*, 2016], which highlights the importance of phenomenal lived-experience from a first-person point of view.

Section 2 of the paper introduces necessary working definitions on the AI side, while Sec. 3 briefly reviews the relationship between phenomenology, enaction, and embodied cognition. Section 4 then proceeds to explore in what manner phenomenal experiences contribute to the cognitive functioning of biological systems, and how they can be of relevance for cognitive systems. In Sec. 5, we explain our proposal to focus on engineering a functional (rather than experiential) equivalent of human phenomenology, where we start our journey towards implementing artificial phenomenology by introducing representationalism as a conceptual foundation bridging cognitive science and AI. Section 6 conceptually contrasts different agentive experiences and discusses attempts to computationally recreate the sense of agency.

We then spell out the requirements for a functional equivalent in cognitive systems and suggest a general blueprint for the implementation of artificial phenomenology. Finally, Sec. 7 summarizes our main arguments and concludes on the potential research- and application-sided impact of successfully engineering artificial phenomenology.

2. Artificial Intelligence and Cognitive Systems

We adopt the description by Nilsson [2009] regarding the aims and means of AI: AI is that science devoted to making machines intelligent, and intelligence is that quality that enables an entity to function appropriately, i.e., to act effectively and with foresight in its environment. This definition allows for a continuum of capacity levels in AI systems, ranging from simple technological systems to human-level machine intelligence [McCarthy, 2007; Besold and Schmid, 2016].

A cognitive system is an AI system that can be considered to be on par with humans in that it is similarly able to perform at least one type of task that corresponds to the exercise of a human cognitive capacity such as goal reasoning, perceiving and responding to different types of stimuli from the environment, or analogy-making. Furthermore, we take a functionalist stance [Piccinini, 2010] in that cognitive systems do not have to confine themselves to methods that are strictly biologically plausible but can employ any technologically realizable means of (re) creating human-level cognitive capacities.

3. Phenomenology, Enaction, and Embodied Cognition

Phenomenology, as a philosophical stance on the nature of reality, can be loosely viewed as a reconciliation of philosophies of idealism and realism [Vernon and Furlong, 2007]. While this relationship is complex, in the following we focus on how phenomenology captures the co-dependence of a cognitive agent and the world in which it is embedded and, consequently, the importance of phenomenal experience to that cognitive agent. Idealism holds that reality is ultimately dependent on the mind of an observer and has no independent existence. In contrast, realism holds that reality exists absolutely and independently of an observer and, either by reason or by sensing, an agent comes to understand its form and structure. Phenomenology offers a third way. According to the phenomenological perspective, our perceptions of the world are a function of what we are: reality is conditioned by experience and experience is conditioned by the nature of the system and its history of interaction with reality. This dependence of reality on the ontogenetic state of an individual is sometimes referred to as radical constructivism [Von Glasersfeld, 2013].

Phenomenology provided the foundation for a new branch of cognitive science — one that takes issue with the cognitivist approach that grew out of the realist tradition [Froese and Ziemke, 2009; Varela, 1992] — known as enaction or enactivism [Varela *et al.*, 2016; Stewart *et al.*, 2010]. It asserts that cognition is a process whereby the issues that are important for the continued existence of a cognitive entity are brought forth or enacted: co-determined by the entity as it interacts with the environment in which it is embedded. Co-determination implies that the cognitive agent is specified by its environment and at the same time that the cognitive process determines what is real or meaningful for the agent. Enaction, in turn, provides a core foundation for embodied cognition, which asserts that "many features of cognition are embodied in that they are deeply dependent upon characteristics of the physical body of an agent, such that the agents beyond-the-brain body plays a significant causal role, or physically constitutive role, in that agents cognitive processing" [Wilson and Foglia, 2017]. Embodied cognition and enactive AI assert the importance of the totality of embodied experience in understanding the world in which the cognitive agent is embedded, including the affective emotional aspects of cognition [Stapleton, 2013]. Sharkey and Ziemke [2001] correspondingly refer to weak phenomenal embodiment where the principles of phenomenal embodiment are simulated by an artificial agent.

The co-determination aspect of our enactive, embodied stance does not make it a solipsist or idealist position of ungrounded subjectivism, but neither is it the commonly-held position of unique — representable — realism. It is fundamentally a phenomenological position. It is worth noting that this does not exclude a representationalist view. However, since we are advocating the importance of experiences and are building on the foundations of phenomenology, the constructivist concept of system-relative representation applies (in contradistinction to classical referential representation): representations derive from "situated cognitive processes whose dynamics are merely modulated by their environment rather than being instructed and determined by it" [Peschl and Riegler, 1999].

4. A potential Role for Phenomenology in Cognitive Systems

A core challenge for every cognitive system is how to best — or, at least, effectively — evaluate its state in and interact with the environment in which it is situated. For humans, there are at least two possible ways of solving these interconnected problems: one route builds upon high-level knowledge-based reasoning capacities, another one relies on phenomenal experiences and the learned understanding of their implications.

The first approach requires perception, representation, reasoning, and evaluation (i.e., the *reasoning route*). Schematically, the pathway from perceiving a property X of the world in a given situation to assessing the valence of this instance of X could go as follows:

(1) Perceive sensory input(s) $\{X\}$.
(2) Represent the perceived inputs: $R_1(\{X\})$.
(3) Provide $R_1(\{X\})$ as input to a high-level reasoning process P, together with information regarding additional external factors F, system knowledge K, etc.: $P(R_1(\{X\}, F, K, \ldots))$.

If terminating successfully, P returns as output a category label for $R_1(\{X\})$, such as pleasure or pain (including the empty label $\emptyset$, i.e., no perceived quality): $P(R_1(\{X\}, \ldots)) {\mapsto} \{\mathit{pleasure}, \mathit{pain}, \ldots \emptyset\}$.

(4) If a category label other than $\emptyset$ has been assigned, a subsequent reasoning step Q then determines an experiential quality in terms of, e.g., weak or strong, based on $R_1(\{X\})$ and the category label: $Q(R_1(\{X\}, \mathit{category})) {\mapsto} \{\mathit{weak}, \mathit{strong}, \ldots\}$.

(5) In most cases, a final evaluative step involving another reasoning process V then uses $R_1(\{X\})$, the (non-empty) category label, and the experiential quality to infer a valence evaluation of this particular instance of X in terms of, e.g., attractive or aversive: $V(R_1(\{X\}, \mathit{category}, \mathit{quality})) {\mapsto} \{\mathit{attractive}, \mathit{aversive}, \ldots\}$. The outcome of this evaluation can, for instance, feed into an action loop, triggering an action that leads to a change in behavior.

Phenomenal experiences by contrast provide immediate, more unmediated access to and evaluation of the possible options open to the cognitive agent. The phenomenological process from perception to evaluation (i.e., the *experiential route*) does not involve high-level knowledge-based reasoning:

(1) Perceive sensory input(s) $\{Y\}$.

(2) Represent the perceived inputs: $R_2(\{Y\})$.

(3) Map from $R_2(\{Y\})$ — and possibly system-internal information S — to an evaluation of the experiential category, quality, and valence in terms of, e.g., pain or pleasure, weak or strong, attractive or aversive: $E(R_2(\{Y\}), S) {\mapsto} \{\{\mathit{pleasure}, \mathit{pain}, \ldots\} \times \{\mathit{weak}, \mathit{strong}, \ldots\} \times \{\mathit{attractive}, \mathit{aversive}, \ldots\}, \emptyset\}$. Again, the result of this mapping might feed into an action loop.

Note that at this point no commitment regarding the precise form of representation appearing in the reasoning route or the experiential route, respectively, has been made. It may well be that distinct representations occur in the corresponding "assign a representation to" steps, in which case R_1 and R_2 may be different in nature and depend on whether the representation subsequently serves as basis for reasoning or for a phenomenal experience.

One notices a similarity between the reasoning route and the common way in which AI in practice conceives of the interface between a system and its environment via evaluative functions. Generally, these functions take into account at least two types/sets of input, either explicitly or implicitly in how the function has been crafted: (i) a set of current system and world states, often together with representations of potential actions of the system, and (ii) a set of goals (i.e., desired system or world states). The function output is an evaluation of the set of states relative to the set of goals. This abstract characterization is applicable independent of the nature of the precise formalism or modeling technique used. But does this conformity with a general approach in AI also mean the reasoning route is the preferable one?

Comparing both approaches, three advantages of the experiential route become apparent: (i) increased efficiency and tractability, (ii) reduced requirements regarding additional information, and (iii) increased generality. Mapping directly from perceptual representations to evaluations removes the reasoning process from representation to category label (step 3 in the schematic outline). This process will often involve the exploration of a significantly-sized state space — due to its reliance on other external information, system knowledge, etc. — or the execution of a lengthy chain of individual reasoning steps, putting a (oftentimes too) heavy performance burden on the reasoner. Moreover, the successful performance of the high-level reasoning mechanism in many cases requires further knowledge, which might not be available to the cognizer at the relevant point in time. Factors constraining access to relevant information can again be performance-related due to, for example, limited computational resources, or be caused by a genuine lack of knowledge from the systems point of view. Phenomenal experiences, by contrast, are assumed to be mostly independent from a cognizers high-level knowledge [Deroy, 2013; Raftopoulos and Müller, 2006]. Finally, generating evaluation functions as required for the reasoning approach is far from trivial and hitherto lacks a general answer or methodology.

Currently, two approaches for creating evaluation functions are in use: Either the system designer directly defines (i.e., *hard-codes*) the functions or she creates a system mechanism generating them following certain rules and patterns (i.e., *learning* them). Both approaches rely on high-level reasoning over processed perceptual input (interpreting perceptual representations with respect to goal states, encoding corresponding evaluation mappings from representations to output values) either *a priori* by the designer or by the system during run-time. This often causes a lack of generality and generalizability because evaluation functions must be grounded in a specific domain or action space so that they can be defined in a comprehensive way. Furthermore, they rely on the presence or absence of defined domain elements or action possibilities, limiting the systems application domains in practice. Here, again, we believe that a phenomenology-inspired approach offers a remedy, since it relies only on the immediate sensory readings of the system and a mapping from sensor outputs and internal system information to experiential evaluation and consequential action, possibly mediated by affective factors [Shanahan, 2006; Ziemke and Lowe, 2009].

We will return to these observations from an implementation-oriented perspective in Sec. 5. For now, they motivate our argument that a computational recreation of phenomenology promises to mitigate several long-standing hindrances on the way towards building AI systems with human-level capacities. This raises the question how one can engineer phenomenology and whether this necessarily means imbuing artificial systems with the ability to perceive phenomenal experiences identical to human phenomenology.

5. Engineering Artificial Phenomenology

We argue for engineering *artificial phenomenology* (i.e., a functional equivalent of phenomenal experiences) rather than human-like phenomenal experiences. The reason for this is two-fold. On the one hand, scientists currently do not have sufficient understanding of how phenomenal experiences arise. In philosophy, this lack of understanding (and the pessimistic outlook on whether understanding will ever arise) is referred to as *the Hard Problem* [Chalmers, 1995]. Although various philosophers and researchers deny the severity of the Hard Problem [Dehaene, 2014; Dennett, 2000], an account of the processes and mechanisms underlying phenomenal experiences is as-of-yet absent. This practically means that there is no starting point for implementation. On the other hand, due to a lack of kinship between artificial systems and humans assuming similarity of the phenomenal experience is unwarranted: It might well be that human phenomenal qualities are an epiphenomenon resulting from the precise forms of representation and/or processing in humans [Dehaene *et al.*, 2018].

At first sight, these might seem like knockout arguments against artificial phenomenology. However, do the phenomenal experiences need to be identical or would a functional equivalent on the side of the machine suffice? Against that backdrop we suggest to focus on engineering a capacity that fulfills the same functions as phenomenal experiences do within cognitive processes but which remains agnostic regarding the actual qualitative dimension. Indeed, one of the principles of phenomenology and enaction is that an agents understanding of its world, and its learned ability to successfully negotiate the difficulties with which it is confronted in its world, is a consequence of its own particular embodiment, including its phenomenal experiences. Consequently, we make recourse to phenomenological research in cognitive science and philosophy where a representational view is often applied to both cognitive capacities as well as phenomenal experiences. In representationalist views of cognition four important elements can be distinguished: The user who uses the representation to guide her behavior, the object (which can be an action or event) that is being represented, the vehicle of the representation which is the physical carrier, and the content, namely, the information carried by the vehicle [Cummins, 1989; Dretske, 1988]. These contents can take different forms; they may be available or unavailable for verbal report, they may involve different senses, etc., [Chalmers *et al.*, 2004]. The basic idea is that an experience is characterized by how the user construes the world. Experiential states can thus be distinguished from each other based on the way the user is representing the world at any given moment.

Representational accounts of experiential states offer a natural interface to approaches in AI falling under the computational cognition paradigm [Pylyshyn, 1980]. To perform computations, a system has to represent the relevant information, independent of the precise form of the corresponding representations. The assumption that cognitive representations are necessary for phenomenal experiences and that the quality of the experience is determined by the content of these representations thus seems natural from the AI point of view — as does the fact that different cognitive systems with different architectural properties may obtain different qualitative experiences from the same representation.

In the remainder of this paper, we use the sense of agency as exemplary experiential phenomenon to spell out the issues that arise when trying to implement artificial phenomenology and our proposed solutions to these challenges.

6. Implementing an Artificial Sense of Agency

Typically developed human adults experience a "sense of agency", i.e., the feeling that one can cause effects through ones actions [Haggard and Chambon, 2012]. The sense of agency contributes to important aspects of human cognition, such as causal learning (e.g., by allowing to learn through intervention [Lagnado and Sloman, 2002]), social and moral interaction (e.g., through responsibility [Caspar *et al.*, 2016, 2018]) and self-awareness (e.g., by being able to distinguish self from other [Jeannerod, 2004; Tsakiris *et al.*, 2007]). The availability of a functional equivalent of the human sense of agency would decisively contribute to cognitive systems engineering.

At least two different phenomena are associated with the term "sense of agency": the "judgment of agency" and the "feeling of agency" [Synofzik *et al.*, 2008]. Upon closer examination, there is a fundamental epistemological difference between both notions. In the case of the judgment, a postdictive reasoning step gives rise to the assumed status as agent in the world — considering oneself as agent provides the best explanation for the observations from the environment [Synofzik *et al.*, 2008]. The judgment of agency is essentially a *post-hoc* belief about one's agency and one's influence in the external world at a given time. In contrast, in the case of the feeling of agency, agency is experienced as a phenomenal quality based on a representation of what the world is like. This representation is thought to come about through a predictive process about the consequences of one's actions and a comparison with the observed state of the world [Synofzik *et al.*, 2008; Blakemore *et al.*, 1998; Zaadnoordijk *et al.*, 2019]. The feeling of agency is not considered a belief as it does not require conceptual content [Bermúdez and Cahen, 2020; Crane, 1992] making it more comparable to a perceptual state instead (for discussion see [Bermúdez and Cahen, 2020; Crane, 1992; Zaadnoordijk and Bayne, 2020]). This difference has significant impact when considering an implementation: while the judgment of agency is equivalent to a form of *ex post* inference to the best explanation[a]

[a] In practical terms, the system must decide whether a change in its environment is most likely due to its own actions. Implementing the reasoning steps required for the judgment of agency, thus, puts several facets of the Frame Problem [Dennett, 2006] on the agenda. Conclusively deciding which aspects of the perceptual input are relevant for the judgment of agency is likely to be computationally intractable (see McDermott [1987] for computational aspects of the Frame Problem), as is the subsequent reasoning process (see McCarthy [1981]; Ginsberg and Smith [1988] for the qualification and ramification aspects of the Frame Problem). Fortunately, the judgment does not have to be infallible — not least because also humans can err when being asked to judge their agency in settings where an immediate observation is not possible [Wegner, 2002].

(i.e., a form of abductive reasoning [Mooney, 2000; Denecker and Kakas, 2002]), the feeling of agency requires an immediate, prospective approach.

Several groups of researchers attempted to equip artificial systems with this prospective sense of agency or closely related capacities. Pitti *et al.*, endowed a robot with the ability to detect and — within "the here and now" [Pitti *et al.*, 2009] — predict contingencies in sensorimotor networks using a neural network emulating spike timing-dependent synaptic plasticity, with the robot starting to act upon the contingencies. The authors took this as behavioral representation of one of the most basic levels of self-awareness and agency. Schillaci *et al.* [2016] followed an approach relying on sensory attenuation [Blakemore *et al.*, 2000] in order to allow a robot to differentiate between self-produced and externally produced actions, interpreting this capability as a first step towards artificial sense of agency. Further examples include Nagai and Asada [2015]'s work implementing a predictive learning architecture that learns about self-other detection, goal-directed actions and helping behaviors by virtue of learning sensorimotor contingencies, the work on body-ownership by Lanillos and Cheng [2018], and the system created by Hwang *et al.* [2018] emulating basic imitation learning of gesture sequences.

What is common to all these projects is that they focused primarily on contingency detection. But obtaining the agency evaluations necessitates one further step beyond the detection of the contingency between predicted and observed world state [Zaadnoordijk *et al.*, 2019]. Following the pattern for phenomenal experiences laid out in Sec. 4, provided with the perceived world state as sensory input, and the predicted world state within the system-internal information at the current point in time, the detection of an equality relation between both generates a mapping to sense of agency as experiential category. This direct mapping allows for a performance evaluation of the phenomenal quality of sensory input without falling victim to the previously discussed resource and generality constraints to which reasoning-based approaches are necessarily subjected. This, of course, unavoidably triggers the question for the genesis of the required mapping function. Different approaches are imaginable, including learning from observed statistical regularities between internal states, bodily movements and subsequent consequences triggered by stimulus-elicited goal-directed behavior; which would be similar to the hypothesized mechanisms allowing human infants to overcome the challenges of bootstrapping the sense of agency [Zaadnoordijk and Bayne, 2020].

7. Conclusion

The challenge in engineering a functional equivalent of human phenomenal experiences resides in leaving out the qualitative dimension without also stripping away the benefits of having phenomenal experiences discussed in the introduction. As argued throughout this paper, since phenomenal cognition is driven by the systems perceptual experience via its sensory input, a possible solution is a direct mapping of certain sensory ranges combined with a snapshot of system-internal information onto immediate "phenomenal values". At this point, the important property is the finite and known range of both the sensors and the internal representational mechanisms of the system (which from a principled point of view holds for static and learning systems alike). The experiential route neither requires an exhaustive enumeration and interpretation (and, thus, a restriction) of the space of possible perceptual states and their representations, nor does it involve a computationally costly evaluation of the current system and world state relative to any goal states. The reduction of relevant information to the perceptual representations together with system-internal properties and application of a direct mapping to qualitative categories with associated evaluation values therefore increases the tractability of the computational process and the generality of the approach. The subsequent output values serve as direct functional counterparts of human phenomenal experiences, for example triggering evasive reactions if undesirable "pain" is encountered or providing positive reward and consequently motivation to continue an action if desirable "pleasure" arises. This facilitates learning and acting in the world in a generalized and tractable way assigning actions based on their predicted outcomes and assessing actual action outcomes, in the manner anticipated by the simulation hypothesis [Hesslow, 2002, 2012].

We grounded our general argument in the context of artificially implementing the sense of agency as assumed foundation of several higher-level cognitive capacities — including causal learning and social interaction. We zoomed in on the feeling of agency explicating the natural correspondence to the experiential route of state assessment as computationally more tractable option. This enables the agent to smoothly perceive its own agency, differentiate its action effects from others and consequently act appropriately relative to its own goals, the social conventions, and so on.

Our proposal coheres with a recent contribution by Man and Damasio [2019] who argued for implementing feelings and motivations into robots. Although their terminology differs from ours, the basis of their proposal is largely consistent what we have suggested here. However, while they focused on the necessity to implement homeostasis to provide robots with their own goals (e.g., survival) and the material properties of the robot to achieve these goals, this paper aimed to formalize artificial phenomenology and is primarily rooted in cognitive (systems) science rather than biology and material science.

In terms of applications, artificial phenomenology promises to unlock a new qualitative dimension in human–computer interaction (HCI) settings, especially in situations involving collaboration and, hence, the establishment of a theory of mind [Meltzoff, 1995]). Artificial phenomenology would greatly contribute to both system behavior resembling human agents as well as complex user-modeling capacities. Regarding the former, consider for instance Forbus [2016]'s "software social organisms". These computational agents are supposed to integrate seamlessly into everyday contexts and act in a way so that "people should be able to relate to [them] as collaborators, rather than tools" [Forbus, 2016]. This involves behavior that can meaningfully be interpreted by humans, similar to the way we rationalize a pet's actions. The system behavior triggered by artificial phenomenology — if properly attuned, in appearance reproducing actions as evoked by phenomenal experiences in humans — will further afford anthropomorphization beyond current levels, following an approach similar to the employment of anthropomorphism in social robotics [Duffy, 2003]. With respect to the augmented user-modeling capacities resulting from artificial phenomenology, given a user's sensory information, the system can have more immediate and better-informed access to the user's cognitive state provided that the labels ("pain", "pleasure", etc.) are mapped to sensory input ranges in such a way as to sufficiently coincide with the actual phenomenal experiences of a human interaction partner. In this way, identifying the likely phenomenal experiences the user is going through based on the environmental conditions, interpreting observed user behavior or forecasting future user actions becomes significantly easier.

Acknowledgments

This work has in part been supported by the MSCA Individual Fellowship InterPlay, No. #891535, awarded to Lorijn Zaadnoordijk. Parts of the work have been carried out at Alpha Health, Telefónica Innovación Alpha, in Barcelona, Spain; at the Radboud University, Donders Institute for Brain, Cognition and Behavior, in Nijmegen, The Netherlands; and at the Department of Electrical and Computer Engineering, Carnegie Mellon University Africa, in Kigali, Rwanda.

References

Bayne, T. and Montague, M. [2011] Cognitive phenomenology: An introduction, in *Cognitive Phenomenology* (Oxford University Press), pp. viii−34.

Bermúdez, J. and Cahen, A. [2020] Nonconceptual Mental Content, in E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*, Summer 2020 ed. (Metaphysics Research Lab, Stanford University).

Besold, T. R. and Schmid, U. [2016] Why generality is key to human-level artificial intelligence, *Adv. Cogn. Syst.* **4**, 13−24.

Blakemore, S.-J., Wolpert, D. and Frith, C. [2000] Why can't you tickle yourself? *Neuroreport* **11**(11), R11−R16.

Blakemore, S.-J., Wolpert, D. M. and Frith, C. D. [1998] Central cancellation of self-produced tickle sensation, *Nat. Neurosci.* **1**(7), 635−640.

Caspar, E. A., Christensen, J. F., Cleeremans, A. and Haggard, P. [2016] Coercion changes the sense of agency in the human brain, *Curr. Biol.* **26**(5), 585−592.

Caspar, E. A., Cleeremans, A. and Haggard, P. [2018] Only giving orders? an experimental study of the sense of agency when giving or receiving commands, *PloS One* **13**(9), e0204027.

Chalmers, D. *et al.*, [2004] The representational character of experience, *The Future for Philosophy*, pp. 153−181.

Chalmers, D. J. [1995] Facing up to the problem of consciousness, *J. Conscious. Stud.* **2**(3), 200−219.

Chella, A. and Manzotti, R. [2011] Artificial consciousness, in *Perception-Action Cycle: Models, Architectures, and Hardware* (Springer, New York), pp. 637−671.

Crane, T. [1992] The nonconceptual content of experience, in *The Contents of Experience* (Cambridge University Press), pp. 136–157.

Crane, T. [2003] The intentional structure of consciousness, in *Consciousness: New Philosophical Perspectives* (Oxford University Press), pp. 33–56.

Cummins, R. [1989] Meaning and mental representation, *Philos. Q.* **40**, 527–530.

Dehaene, S. [2014] *Consciousness and the Brain: Deciphering how the Brain Codes Our Thoughts* (Penguin).

Dehaene, S., Lau, H. and Kouider, S. [2018] Response to commentaries on what is consciousness, and could machines have it, *Science* **359**(6374), 400–402.

Dehaene, S. and Naccache, L. [2001] Towards a cognitive neuroscience of consciousness: basic evidence and a workspace framework, *Cognition* **79**(1–2), 1–37.

Denecker, M. and Kakas, A. [2002] Abduction in logic programming, in *Computational Logic: Logic Programming and Beyond* (Springer), pp. 402–436.

Dennett, D. [2000] Facing backwards on the problem of consciousness, *Explaining Consciousness — The "Hard Problem"*, pp. 33–36.

Dennett, D. [2006] The frame problem of AI, *Philosophy of Psychology: Contemporary Readings*, Vol. 433, pp. 67–83.

Deroy, O. [2013] Object-sensitivity versus cognitive penetrability of perception, *Philos. Stud.* **162**(1), 87–107.

Dretske, F. [1988] *Explaining Behavior: Reasons in a World of Causes* (MIT Press).

Duffy, B. R. [2003] Anthropomorphism and the social robot, *Robot. Auton. Syst.* **42**(3–4), 177–190.

Forbus, K. D. [2016] Software social organisms: Implications for measuring AI progress, *AI Mag.* **37**(1), 85–90.

Froese, T. and Ziemke, T. [2009] Enactive artificial intelligence: Investigating the systemic organization of life and mind, *Artif. Intell.* **173**(3–4), 466–500.

Gallagher, S. [2004] Understanding interpersonal problems in autism: Interaction theory as an alternative to theory of mind, *Philos. Psychiatry, Psychol.* **11**(3), 199–217.

Gallagher, S. and Zahavi, D. [2013] *The Phenomenological Mind* (Routledge).

Gallese, V., Eagle, M. N. and Migone, P. [2007] Intentional attunement: Mirror neurons and the neural underpinnings of interpersonal relations, *J. Amer. Psychoanal. Assoc.* **55**(1), 131–175.

Gilbert, D. T. and Wilson, T. D. [2007] Prospection: Experiencing the future, *Science* **317** (5843), 1351–1354.

Ginsberg, M. L. and Smith, D. E. [1988] Reasoning about action I: A possible worlds approach, *Artif. Intell.* **35**(2), 165–195.

Haggard, P. and Chambon, V. [2012] Sense of agency, *Curr. Biol.* **22**(10), R390–R392.

Haggard, P. and Clark, S. [2003] Intentional action: Conscious experience and neural prediction, *Conscious. Cogn.* **12**(4), 695–707.

Hesslow, G. [2002] Conscious thought as simulation of behaviour and perception, *Trends Cogn. Sci.* **6**(6), 242–247.

Hesslow, G. [2012] The current status of the simulation theory of cognition, *Brain Res.* **1428**, 71–79.

Hwang, J., Kim, J., Ahmadi, A., Choi, M. and Tani, J. [2018] Dealing with large-scale spatio-temporal patterns in imitative interaction between a robot and a human by using the predictive coding framework, *IEEE Trans. Syst. Man, Cybern., Syst.* 1–14.

Jeannerod, M. [2004] Visual and action cues contribute to the self–other distinction, *Nat. Neurosci.* **7**(5), 422–423.

Keltner, D., Horberg, E. J. and Oveis, C. [2006] Emotions as moral intuitions, *Affect in Social Thinking and Behavior*, pp. 161–175.

Lagnado, D. A. and Sloman, S. [2002] "Learning causal structure," in *Proc. Annual Meeting of the Cognitive Science Society*, pp. 560–565.

Lamme, V. A. [2006] Towards a true neural stance on consciousness, *Trends Cogn. Sci.* **10**(11), 494–501.

Lanillos, P. and Cheng, G. [2018] "Adaptive robot body learning and estimation through predictive coding," in *2018 IEEE/RSJ Int. Conf. Intelligent Robots and Systems (IROS) (IEEE)*, pp. 4083–4090.

Man, K. and Damasio, A. [2019] Homeostasis and soft robotics in the design of feeling machines, *Nat. Mach. Intell.* **1**(10), 446–452.

Mandler, G. [1989] Affect and learning: Causes and consequences of emotional interactions, in *Affect and Mathematical Problem Solving* (Springer), pp. 3–19.

McCarthy, J. [1981] Epistemological problems of artificial intelligence, in *Readings in Artificial Intelligence* (Elsevier), pp. 459–465.

McCarthy, J. [2007] From here to human-level AI, *Artif. Intell.* **171**(18), 1174–1182.

McDermott, D. [1987] We've been framed: Or, why AI is innocent of the frame problem, in *The Robot's Dilemma: The Frame Problem in Artificial Intelligence* (Ablex), pp. 113–122.

Meltzoff, A. N. [1995] Understanding the intentions of others: re-enactment of intended acts by 18-month-old children, *Dev. Psychol.* **31**(5), 838.

Mooney, R. J. [2000] Integrating abduction and induction in machine learning, in *Abduction and Induction: Essays on their Relation and Integration* (Springer), pp. 181–191.

Nagai, Y. and Asada, M. [2015] "Predictive learning of sensorimotor information as a key for cognitive development," in *Proc. IROS 2015 Workshop on Sensorimotor Contingencies for Robotics*.

Nilsson, N. J. [2009] *The Quest for Artificial Intelligence* (Cambridge University Press).

Pacherie, E. [2008] The phenomenology of action: A conceptual framework, *Cognition* **107**(1), 179–217.

Peschl, M. F. and Riegler, A. [1999] Does representation need reality? in *Understanding Representation in the Cognitive Sciences* (Springer), pp. 9–17.

Piccinini, G. [2010] The mind as neural software? understanding functionalism, computationalism, and computational functionalism, *Philos. Phenomenol. Res.* **81**(2), 269–311.

Pitti, A., Mori, H., Kouzuma, S. and Kuniyoshi, Y. [2009] Contingency perception and agency measure in visuo-motor spiking neural networks, *IEEE Trans. Auton. Ment. Dev.* **1**(1), 86–97.

Pylyshyn, Z. W. [1980] Computation and cognition: Issues in the foundations of cognitive science, *Behav. Brain Sci.* **3**(1), 111–132.

Raftopoulos, A. and Müller, V. C. [2006] The phenomenal content of experience, *Mind Lang.* **21**(2), 187–219.

Schillaci, G., Ritter, C.-N., Hafner, V. V. and Lara, B. [2016] Body representations for robot ego-noise modeling and prediction towards the development of a sense of agency in artificial agents, *Artif. Life Conf. Proc. (28)*, 390–397.

Shanahan, M. [2006] A cognitive architecture that combines internal simulation with a global workspace, *Conscious. Cognit.* **15**(2), 433–449.

Sharkey, N. E. and Ziemke, T. [2001] Mechanistic versus phenomenal embodiment: Can robot embodiment lead to strong AI? *Cognit. Syst. Res.* **2**(4), 251–262.

Sloman, A. and Chrisley, R. [2003] Virtual machines and consciousness, *J. Conscious. Stud.* **10**(4–5), 133–172.

Stapleton, M. [2013] Steps to a properly embodied cognitive science, *Cognit. Syst. Res.* **22**, 1–11.

Stewart, J., Gapenne, O. and Di Paolo, E. A. [2010] *Enaction: Toward a New Paradigm for Cognitive Science* (MIT Press).

Synofzik, M., Vosgerau, G. and Newen, A. [2008] Beyond the comparator model: A multifactorial two-step account of agency, *Conscious. Cognit.* **17**(1), 219–239.

Torrance, S. [2008] Ethics and consciousness in artificial agents, *AI & Soc.* **22**(4), 495–521.

Tsakiris, M., Schütz-Bosbach, S. and Gallagher, S. [2007] On agency and body-ownership: Phenomenological and neurocognitive reflections, *Conscious. Cognit.* **16**(3), 645–660.

Varela, F. J. [1992] Whence perceptual meaning? a cartography of current ideas, in *Understanding Origins* (Springer), pp. 235–263.

Varela, F. J., Thompson, E. and Rosch, E. [2016] *The Embodied Mind: Cognitive Science and Human Experience* (MIT press).

Vernon, D. and Furlong, D. [2007] Philosophical foundations of AI, in *50 years of Artificial Intelligence* (Springer), pp. 53–62.

Von Glasersfeld, E. [2013] *Radical Constructivism* (Routledge).

Wegner, D. M. [2002] *The Illusion of Conscious Will* (MIT press).

Wilson, R. A. and Foglia, L. [2017] Embodied Cognition, in E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*, Spring 2017 ed. (Metaphysics Research Lab, Stanford University).

Zaadnoordijk, L. and Bayne, T. [2020] The origins of intentional agency, doi10.31234/osf.io/wa8gb, psyarxiv.com/wa8gb.

Zaadnoordijk, L., Besold, T. R. and Hunnius, S. [2019] A match does not make a sense: On the sufficiency of the comparator model for explaining the sense of agency, *Neurosci. Conscious.* **2019**(1), niz006.

Ziemke, T. and Lowe, R. [2009] On the role of emotion in embodied cognitive architectures: From organisms to robots, *Cognit. Comput.* **1**(1), 104–117.