FILOZOFIA I NAUKA
Studia filozoficzne i interdyscyplinarne
Tom 10, zeszyt specjalny, 2022

Jeffrey White

# ON A POSSIBLE BASIS FOR METAPHYSICAL SELF DEVELOPMENT IN NATURAL AND ARTIFICIAL SYSTEMS

doi: 10.37240/FiN.2022.10.zs.4

***ABSTRACT***

Recent research into the nature of self in artificial and biological systems raises interest in a uniquely determining immutable sense of self, a "metaphysical 'I'" associated with inviolable personal values and moral convictions that remain constant in the face of environmental change, distinguished from an object "me" that changes with its environment. Complementary research portrays processes associated with self as multimodal routines selectively enacted on the basis of contextual cues informing predictive self or world models, with the notion of the constant, pervasive and invariant sense of self associated with a multistable attractor set aiming to ensure personal integrity against threat of disintegrative change. This paper proposes that an immutable sense of self emerges as a global attractor which can be described as a project ideal self-situation embodied in frontal medial processes during more or less normal adolescent development, and that thereafter serves to orient agency in the more or less free development of embodied potentials over the life course in effort to realize project conditions, phenomenally identified with the felt pull towards this end as purpose of and source of meaning in life. So oriented, life-long self-development aims to embody solutions to problems at different timescales depending on this embodied purpose, ultimately in the service of evolutionary processes securing organism populations against threats of disintegrative change over timespans far beyond that of the individual. After characterizing the target sense of self, research circling this target is briefly surveyed. Self as global project and developmental neural correlates are proposed. Then, the paper discusses some implications for research in biological and artificial systems. Building from earlier work in cognitive neurorobotics, discussion affirms the value of reinforcement rituals including prayer in metaphysical self-development, considers implications for value alignment and rights associated with free will in the context of artificial intelligence and robot religion, and concludes by emphasizing the importance of self-development toward project ideals as source of meaning in life in the current social-political environment.

**Keywords**: self, purpose in life, default mode network, predictive processing, AI value alignment, developmental robotics.

## 1. INTRODUCTION

Current predictive coding (PC) and predictive processing (PP) inspired research into self, including that grounded in the tandem principle of free energy and active inference (FEP/AIf), suggests that different phenomena associated with self emerge through ongoing iterative interaction of prospective body schema with the objective world, with discernible senses of self presenting at different timescales as anticipations of possibly perceived conditions of future embodied situations are contravened (Tani, White, 2020; cf. White, Tani, 2016; 2017; Tani, White, 2016; Hohwy, Michael, 2017; Williford et al., 2018; Safron, 2021, for review). However, questions remain concerning an immutable "metaphysical" self distinguishable from more "minimal" senses of self. This paper works at answering these questions.

Mateusz Wozniak (2018) analyses self into object "me" and subject "metaphysical 'I'," locating "me" in hierarchical layers of neurological activity differently affected by changes in the environment. His account burdens researchers maintaining the existence of an "I" remaining constant in the face of environmental change to locate this sense of self in hierarchical structural dynamics, similarly. Jose Ortega y Gasset (OyG) (2002) identifies a pervasive, constant sense of self with a global project ideal developing as a propositional self-model establishing a life-long motivational goal-orienting internal self-relation characterized as a calling forward to one's self in terms of "vocation." This paper proposes that such a self-relation develops more or less normally during adolescence in a value oriented subsystem of the default mode network (DMN) in human beings, and considers that similar developmental dynamics may be formalized for artificial intelligence (AI) applications with implications for robot rights and AI value alignment.

The next section begins with Wozniak's challenge to account for a metaphysical self, and reviews Klaus Gärtner and Robert Clowes' (2020) analysis and counter-proposal. The third section surveys complementary research including Rutger Goekoop and Roy deKleijn's (2021a) "bowtie" model. The fourth section introduces OyG's phenomenological account of self as vocation, and correlates this with adolescent development. The fifth section develops the bowtie into the "traveling bowtie" before illustrating focal structural dynamics with the Platonic cave. The paper considers implications of the present view for AI research in the context of prayer as value-reinforcing ritual, robot religion, value alignment and robot rights in the sixth section, and concludes with critical observation of contemporary challenges for self-development of prosocial project ideals in the seventh.

## 2. METAPHYSICAL SELF

Wozniak (2018) analyzes self in the context of predictive coding (PC) including Friston's FEP/AIf.[1] He distinguishes "I" from "me" via Ludwig Wittgenstein's (1958) illustration of an "I" seeing a "me" in a mirror, i.e. "I" see "me." And, he reduces talk of different senses of self—including the "intuitive understanding of subject-of-experience as continuously persisting lifelong stream of consciousness" (Wozniak, 2018, p. 9) that characterizes the "metaphysical 'I'"—to instances of the object "me" sense of self emerging as specific aspects of hierarchical neural structures are affected by changes in perceived reality over different timescales in different, increasingly integrated sensory streams.

Wozniak recognizes that PC inspired accounts help to clarify structural dynamics responsible for different senses of self. On such accounts, layers of a hierarchy generate predictive models of causes of input from lower layers. Sensory states are associated with error between predictions and perceived reality as this information is passed upwards and the model hierarchy is updated in the direction of minimizing subsequent error. Actions undertaken from updated system states aim to change the order of the object environment (and, thereby the internal model) in the same direction. Environmental changes are subsequently perceived, serving as input in the next time-step in continuous circular causality between an agent's internal predictive models (e.g. prior beliefs, anticipations, projections, actions) and that agent's environment.

In this context, Wozniak argues that any sense of "I" is best understood as a particular sense of "me" because dynamics responsible for the phenomenon should arise between levels of activity in a temporal hierarchy, just as does the sense of "me," if only in the form of a "delusion" maintained regardless of environmental change. Wozniak then presses the question about where a constant sense of self associated with the "metaphysical 'I'" may be found in a natural system, so understood. If not located as phenomena associated with "me" in layers of a temporal hierarchy, then this sense of "I" may be deflated away from technical discourse, leaving only those senses which *can* be located in nature in either direction of the perception-action stream, i.e. self-as-object "me" perhaps in network dynamics manifesting as a delusion of a constant and immutable subject "I" motivating change-ignorant

[1] Wozniak uses "predictive coding" while others use "predictive processing" to discuss the same sorts of structural dynamics. Though distinguishable, literature closer to cognitive robotics and systems programming often shows predictive coding and that closer to cognitive science often predictive processing, though other distinctions might be possible i.e. perhaps using PP when emphasizing forward processing, e.g. active inference, with PC about message passing upward as in the perceptual or bottom-up open mode (of the ACTWith model, for example). This paper uses the terms as do the represented authors, treating them as a family of accounts focused on prediction and error minimization in temporal hierarchical structures whether in natural or artificial systems, with the outstanding question being to what end.

and context-inappropriate actions such as holding out for such a sense of self. In the end, Wozniak challenges those who wish to maintain the reality of an immutable subject "I" (that is not in reality only a deluded "me") to "prove that there is a qualitative difference between them, and to demarcate the exact border." (Wozniak, 2018, p. 12) In the absence of such a proof, Wozniak suggests that PC and related approaches "can attempt to retain relevance" by inquiring into Ned Block's (1995) "access consciousness" characterized as a "functional mechanism" allowing for "attended information" to enter awareness and become reportable to others (paraphrasing from p. 11, with Wozniak quoting Dehaene, 2014). This paper returns to access to and communication of a sense of self answering to Wozniak's challenge in section 4.[2]

Wozniak recognizes an account of a pervasive and constant sense of self within PC associated constraints in Thomas Metzinger's self-model theory (Metzinger, 2009). Metzinger asks "Is there a fundamental (and perhaps implicit) kind of phenomenal character sui generis, which can at times be made explicit and which underlies or "permeates" all other forms of phenomenal experience?" (Metzinger, 2020, p. 6) In answer, he builds an account of a "primordial" form of "pure consciousness," "pure awareness" and "bare wakefulness" as "minimal phenomenal experience" (MPE) which is "aperspectivial," of an "indivisible [...] epistemic space" "as yet without object" and without a sense of "self-location in a spatial frame of reference" (p. 37). Metzinger's MPE is essentially "non-egoic" without "self-location in time" or "space" and without "quality of agency" (p. 10), yet it grounds self-experience, being the "natural state" of an agent "predicting itself into existence" (note 26, p. 38, quoting Friston) from the potential of which minimal phenomenal self (MPS) arises with corresponding senses of self-location, perspective and purpose (cf. Williford et al., 2018). I think that Metzinger is wrong about MPE being a self-predicting agent's "natural state" and offer a correction culminating in the conclusion to this paper.

In PC inspired computational models, Wozniak's "me" may be associated with senses beginning with "minimal self" at lowest and most immediate timescales, with increasingly abstract conceptions including "narrative self" associated with activity in higher layers of the temporal hierarchy as primitive instances are integrated into larger patterns of episodic activity over longer timescales (see Tani and White, 2020). Here, it is worth noting that the "metaphysical 'I'" corresponding with the subject that sees its object self in the mirror could correspond with the top layer of activity in recurrent neural networks constrained by different timescales of processing at different levels in different modalities as these are integrated upwards, with high-

[2] Without a corresponding sense of self on the other end, however, communication may be practically impossible. Mired in the philosophy department at University Twente for instance, I found myself saying: "You can't see it if you can't see it."

er order processes modeling increasing invariance associated with context independence and constancy such as in the case of moral principles and their exemplars. Top-level activity generally is characterized as "intentional" being the final layer corrected given error as prior intentions are enacted and misalignments with perceptual reality mediated through iterative interaction with the object environment (cf. Tani, 2017; Limanowski, Friston, 2020). Goekoop and DeKlein (G&dK) (2021a) characterize such structural dynamics in terms of throughput layers integrating input and output information streams using the image of a "bowtie" and extend this basic model to interpersonal and social dynamics (central to G&dK, 2021b). More is made of these ideas in the next section.

## 3. COMPLEMENTARY VIEWS

Gärtner and Clowes (G&C) (2020) also assess metaphysical self in the context of Wozniak (2018). On their analysis, using the term predictive processing (PP), such accounts are constrained along two dimensions, one being that self changes as affected by environment, the "mutability" constraint, and the other being that self is multi-layered. Due to these constraints and consistent with Wozniak, on their view PP approaches have difficulty accounting for metaphysical self, supporting what they call "anti-realism" about self, due primarily to the mutability constraint. The present paper offers a PP inspired realist account in section 4.

G&C introduce their "pre-reflective situational self" as a possible account of metaphysical self (Clowes and Gärtner, 2020). On this model, Wozniak's "I" corresponds with a collection of "situational self positions" according to which an agent acts more or less appropriately ("pre-reflectively") in different (including specifically social) contexts. G&C's situational self involves multi-layered processing from pre-reflective to reflective consciousness comparing intentions as possible situations that are determined by and change according to situational demands, in short representing a standard PP account while also accommodating "relational" views in terms of which selves exist in the context of other selves, socially, with each individual occupying a unique position that is essentially (i.e. informed by the embodied mirror system) relative to others. Importantly thus, their situated self is essentially normative as an agent "fluidly and appropriately" adapts "spontaneously and naturally in the context of managing everyday life" (p. 72; compare Limanowski and Friston's "transparent" per discussion below) including while navigating social norms and expectations of others sharing in and contributing to the embodied situation. As the social organism shifts between different contextually dependent roles, its "self-positions" can be thought of as embodied sub-routines associated with feelings, attitudes and

emotions more or less appropriate for a given situation, with shifts between sub-routines proceeding unconsciously according to operational context, "pre-reflective," and with the repertoire as a whole associated with metaphysical self.

Cognizant of Metzinger's non-realist "no-self" view, G&C argue that theirs is a "realist" account in which the self is a substantial, "constant entity" and "labile aspect of the phenomenal field which while changing continues to play the same role and, very importantly, occupies the same place" (Gärtner, Clowes, 2020, p. 73; cf. Newen, 2018). Self is not experienced as a "stable and unchanging subject" but is perceived as mutable, emerging in different ways in context-dependent error-passing upwards through layers of processing, and can be identified with these events as is Wozniak's "me." At the same time, G&C argue that their situational self is constant as ongoing adaptation to situational constraints is essential to uniquely embodied self-perception (cf. Valmisa's, 2021, treatment of situations, similarly). Thusly, G&C take the uniquely embodied situation and associated phenomena to be fundamental to self and so constant, rather than filler to be abstracted away from a formal envelope as does Metzinger.

Dynamic and multilayered, mediated by context-dependent behavioral repertoires more or less skillfully enacted, G&C's view resembles the multistable attunement of ecological enactivism (cf. Bruineberg et al, 2021) for which neurological grounds can be discovered in "ghost attractors" embodied in DMN dynamics (cf. Deco, Jirsa, 2012). And, G&C (2020) survey a number of alternative PP inspired accounts which paint a similar portrait, including that of Chris Letheby and Phillip Gerrans (2017) who account for self in terms of binding across systems as attention and corresponding context-dependent phenomenal contents change. Likewise, G&C review Wanja Wiese's (2019) "SANTA" model accounting for persistent sense of self in terms of attentional shifts that are accompanied by a pervasive feeling of control over ongoing actions (there is a lot of recent attention to this idea in different areas, e.g. Sennesh et al., 2022, in the context of predictive processing and interoception as allostasis; Kahl et al., 2021, in the context of artificial systems; foundationally, see Sterling, 2012). And, on this model, self becomes evident at highest levels of contextually dependent processing.

Jakub Limanowski and Karl Friston (L&F) also locate self at the highest levels of processing (Limanowski, Friston, 2020). L&F write that "the 'self'" is "a hypothesis or latent state (of being) that can be associated with a self-model" that "arises as (computationally) the most accurate and parsimonius explanation for bottom-up multi-sensory information" (ibidem, p. 3) realized through action in differences between expectation and perceived reality. At the same time, L&F recognize that self involves a special case of active inference that is inward, interoceptive, whereby an agent may act on itself, adjusting internal structural dynamics in order to satisfy goal-directed in-

tentions in the overall aim of minimizing free energy, e.g. modulating anxieties about uncertainties through meditation. Complementarily, Sennesh et al. (2022) discuss such activity in terms of allostatic control.

L&F (2018) propose an account of self that answers to Wozniak's "I" as transparent intentions guiding actions according to top-down predictions. Following Metzinger's "self-model theory" (Metzinger, 2003), a self-model becomes a phenomenal self-model as intention fails to deliver to anticipation. The basic idea is already familiar, that intentions are enacted top-down through timescales in effort to coordinate with focal aspects of the object environment through action toward situations with reduced uncertainty and with it potential for integrity threatening surprise. Upwards through the hierarchy, phenomenal contents, including "representations" for introspected attention, manipulation and communication, become increasingly invariant in the face of environmental change, with the "reality" of an object, whether material or in the form of a delusion per Wozniak, corresponding with this model invariance. So on this account, an agent becomes aware of its self as an "epistemic agent" as it exercises a capacity to selectively attend to different features of the perceptual stream, and moreover to actively construct action plans and manipulate mathematical forms ("representations") through "introspective attention" exercising "epistemic agency" over a "representational space" (drawing from Blank, Metzinger, 2009; cf. Wiese's "salience object"). Reminiscent of accounts surveyed above, self as an invariant concept corresponds with that bundle of routines by way of which an agent adjusts to the changing world (perhaps as self and world models develop in parallel per Newen, 2018), summarily in order to maintain embodied integrity (of this bundle) in the face of disintegrative change. Like G&C's situational self, this multi-stable capacity for selective attention is constant, and as with Metzinger's envelope, once self-phenomena are abstracted away, describes something necessary for any experience at all (see Pezzulo et al., 2021, for interesting parallels with this envelope structure, as well).

L&F (2020) offer an interpretation of "selfless" experience of the sort from which Metzinger's view emerges—in terms of which self-experience emerges from something inaccessible to introspection. Following Metzinger (2003) an organism proceeds mostly unaware, with "self-models" "transparent" and not present as objects of attention. Naturally, this inclination to routine makes sense, as introspected attention (and higher order thought generally speaking) is computationally and metabolically costly. Accordingly, the structural hierarchy of FEP inspired approaches involves the reduction of complexity and attention-demanding activities into routine operations in order to reduce metabolic demands, thereby freeing up higher-order capacities to attend to outstanding concerns or to rest in transparent enaction of learned priors. So, L&F argue that the transparent state is the basic one, consistent with Metzinger's, G&C's and related accounts, and that what

is necessary is a constraint on attention in order to keep metabolic costs to a minimum in the ongoing refinement of enacted routines which they discuss in terms of "precision." Precision involves ever-finer-grained determination of world as navigated and self as embodied internal dynamics are revealed through iterative interaction (again, compare Newen, 2018). This includes the social world, and so they offer what is in effect an account of attunement of higher-order processes in development of context-dependent including social-normative sub-routines as in G&C's account and as reflected in contemporary enactivist literature. The present paper accounts for necessary constraints on attention beginning with Ortega y Gasset's "vocation" in the next section.

Consider in this context Goekoop and deKleijne's (2021a) "bowtie" model, with input consisting of multi-modal streams fed upward through a temporal hierarchy established as these streams converge and are integrated with complexity proportional to the "independent contextual cues that need to be controlled by the organism," "throughput" layers (the knot of the bowtie) which bridge input to output streams at higher levels of these processes characterized in terms of "width" of "bottlenecks" associated with intentions as discussed above (p. 264; section 3.3, box 1, p. 263 details the "bowtie hierarchy"), and output streams which feed intentions down the hierarchy in actional coordination with the object environment. G&dK argue that bow-tie structures spontaneously emerge under evolutionary constraints of scarce resources e.g. food, time, satisfying needs by "compressing" necessary operations into actionable intentions e.g. how to get the most food in the least time, which ostensibly may be communicated as a series of steps and/or set of guiding principles (cf. Nyberg et al., 2022, for interesting corollary at the level of goal-related memory).

G&dK link the life-long "outgrowth and sculpting" of bowtie structure "goal hierarchies" with "personality development" as organisms mature through "different forms of associative learning … in relation to themselves and their environments" (2021a, p. 276). Roughly, the view offered here is that goal-hierarchies mature at three levels of functionality—self-referential (perhaps associated with self as an active situation, cf. Valmisa, 2021), intersubjective or social (perhaps associated with self relative others as mutual input streams), and normative (perhaps associated with relative invariance of principles and moral exemplars as self models)—over "a life-long process of goal-directed learning" i.e. "personality development" (G&dK, 2021a, p. 276). High-level processes embody "global states" which "harbor some of the most global ('domain general') representations of the inner and outer environment" (i.e. self and world models) and which "bias activity levels in several subordinate brain areas involved in the planning and execution of motor programs, which control a multitude of pyramidal cells and muscle fibers to produce motor action" (G&dK, 2021a, p. 262; here, following L&F,

we may consider that the learning system aims to increase precision while minimizing path length according to fundamental physical principles via bowtie throughput layers). As with L&F's invariance, G&dK identify most-connected (highest organizational level) nodes with "social norms and moral values that individuals deem applicable across living systems and time-scales" (G&dK, 2021a, p. 277). Such norms and values can be associated with injunctions not to harm, not to lie, not to use others as a means for one's own ends, with the stress-induced (perhaps due to someone lying, causing harm, and misleading for personal enrichment at the expense of others) incapacity to continue in principled goal-seeking causative of "moral decay" in selves and social systems thereby affected.

The focus of G&dK (2021a) is to account for the effects of stress on high-level processes, with excess chronic stress causing mental and personality disorders. The central idea is that higher-level processing is neglected as stress constrains attention to more immediate conditions. With stress, "error accumulates vertically in the goal hierarchy and increases the oscillation frequency of network nodes until energy demand exceeds energy supply ('allostatic overload')" (G&dK, 2021a, p. 276) resulting in metabolic incapacity to retain higher-level goals. The "most connected" (in the sense of small world dynamics) "nodes at the top of the goal hierarchy are most vulnerable to such energy depletion, causing them to selectively overload and fail" with relevant dynamics "undercontrolled" (G&dK, 2021a, p. 276). Mental and personality "disorders" are evidenced in the "collapse of goal hierarchies" as lower-level demands make higher-level processing impossible, with more "strongly matured" hierarchical structures better able to "withstand the pruning of their hierarchies during a stressful episode" (G&dK, 2021a, p. 276). Briefly, we may picture the throughput layer of the bowtie moving up and down the hierarchical structure in the service of stress reduction through action according to contextual demands. Mental and personality disorders present as incapacities to shift across operational contexts and so to adjust throughput processing in appropriate ways, perhaps resulting in persistent self-phenomena e.g. Wozniak's "delusions".

G&dK distinguish between personality and mental disorders according to how they develop. Mental disorders involve "temporary" dissolution of "high-level (integrative) goal states ... e.g. major depression, psychosis, panic attacks)" while "personality (trait) disorders" or "personality deficits" involve a failure of goal hierarchies to develop normally and to "mature in the course of life" (G&dK, 2021a, p. 277). On the relationship between stress and different disorders, they point to neuroimaging studies demonstrating reduced grey matter volume in the same areas of human brains down-regulated during stressful episodes, with symptoms including "decreased sense of purpose" and involving under-developed "normative functions" as well as "self-referential" and "intersubjective" functions and with, in "(bor-

derline) personality disorder", "underdeveloped brain areas" involving "the same areas that harbor our world models of self, others and global world views" accordingly (G&dK, 2021a, p. 276). In the context of their overall view, they note that the word "disorder" is "well-chosen" as sensitivity to certain stimuli potentiates responses which, through circular causality with the triggering environment, "signal a loss of homeotasis" leading to "disease" and "death" and with such dynamics extending to "any scale level of organization, including social levels." (G&dK, 2021a, p. 277; cf. G&dK, 2021b)

G&dK (2021a) point to the promise of research into especially pathological interpersonal dynamics emergent in terms of cascades of input-output loops as "undercontrolled (stressed) individuals" develop strong co-dependencies potentiating "a mutual loss of law-abiding and moral behavior" (e.g. Bonny and Clyde, a home-robbing street gang). On their account, higher levels of social organization including social network clusters demonstrate emergent ingroup-outgoup dynamics and in so doing "may follow similar rules for network architecture and function (collective inference) as shown in hierarchically organized input (perception), throughput (goal setting) and output (action) parts that are engaged in Bayesian inference" (p. 276; cf. G&dK, 2021b). As "vicious cycles in social behavior" emerge due to "insufficient higher-level control" and "typically require an external party" to interrupt destructive feedback loops, such studies might constructively inform social policy (G&dK, 2021a, p. 276). G&dK thus extend the basic bowtie model optimizing throughput to group dynamics in which individual output serves as input for others. In the case of borderline personalities, for example, the general thesis is that stress during critical developmental periods affects embodied network structure subsequently modulating behavior during stressful periods, which then serves as more or less disordering input for surrounding bowtie systems, resulting in cascading dysfunction at higher levels of organization by way of a mechanisms which may be considered in terms of "resonating minds" (as described by Poppel et al., 2021) as higher-level processes anticipate goal-hierarchical collapse and act accordingly, thereby establishing potentially dysfunctional norms at higher levels of social-political organization, presumably extending to mass psychosis and hysteria (cf. Bagus et al., 2021).

Situated in co-evolutionary time scales with goal hierarchy maturation tempered by cultural and historical constraints, G&dK's view comes closer to establishing a constant sense of self within a PC consistent framework, one that resists disordering influences. However, holding out for a constant sense of self as a sort of highest-order immutable goal-state, in the face of contextual especially normative demands, would seem to invite charges of personality disorder as the embodied bowtie struggles to maintain such goals against normative stressors, resulting in erratic behavior and so

apparent dysfunction including norm violation perhaps perceived as immoral. "Maturity" thus might involve letting go of for example deeply principled self-associations, foregoing pursuit of a moral exemplar in resonant attunement to more immediate social expectations.

It is not clear how and when highest-order embodied goals should be foregone due to interests in personal safety, saving others the stress of not "going along to get along" to "fit in" perhaps while risking an expert "borderline personality" diagnosis, especially when trying to account for "evolutionary goals" that presumably are not constrained by current cultural-historical standards, as do G&dK. Why should an agent attuned to situational constraints at evolutionary time-scales give up on these goals, perhaps working to ensure not only the survival of but flourishing future humanity, when confronted by a contemporary political economy which rewards behavior to the contrary, encouraging the exchange of highest-order goals for fiat currencies and material luxuries simply in order to minimize stress for passing personal well-being? Wouldn't the morally principled thing be to maintain those aspirations somehow, suffering the dissolution of lower-level goals including perhaps bodily integrity through unjust punishments and loss of in-group support of contemporaries, instead?[3] This is not clear on G&dK's account, the line between higher-order and disorder. What is missing is an account of the retention of higher-order goals in the face of more immediate pressures, ideally in the form of a mechanism underwriting motivation to order contrary to established norms, that both resists dissolution and that is not also evidence of personality disorder or self-delusion. Such an account is proposed in the next section.

## 4. SELF PROPOSAL

How might a "metaphysical 'I'" that is not reducible to Wozniak's "me" and that is not constrained to evident norm satisfaction arise in a temporal hierarchy such as those discussed so far in this paper, perhaps formalized for applications in the context of developmental robotics and AI? Discussion left off with bowtie hierarchical goal structures mediating the perception-action loop through compressed higher-level intentional layers embodying goals relatively detached from and invariant to environmental change, and

[3] Directly contra Miller et al.'s (2021) recommendation to relax "rigid" associations for long-term "well-being" optimizing for "happiness" in the near-term, note that the present paper works from a teleological understanding of happiness, Aristotelian purposeful rather than pleasant, reinforcing the point that trading principle for personal security may not be of significant value. Summarily, where Miller et al. propose that agents seek slopes for informative error reduction via externally sourced "affordances" in potential self-realization, their account reflects dynamics associated with posterior DMN dynamics but neglects the internal slope corresponding with the metaphysical self as global attractor with corresponding affordances "self-affordances" associated with anterior DMN (particularly dorsal medial) dynamics as developed in the present paper.

with development and ongoing refinement of subservient throughput operations associated with personality development. Whether rendered in terms of enactivist skillful attunement, shortcut throughput layers in human beings or predictive codes passing messages down through computational hierarchies in biologically inspired neurorobots, such PC inspired models minimize error of fit to environment as agents "attune" themselves through enacted prior embodied anticipation in the perceptive enhancement of control over internal (embodied) and external environment (together, G&C's "situation"). Exercised in the reduction of disease and death inducing stress evidenced in the dissolution of higher-order goals due allostatic overload per G&dK (2021a), self is revealed in the breakdown (in robots, see Tani, 2017, on self-organized criticality and minimal self). G&dK associate resistance to "pruning" of such higher-order processes with "maturity" of bowtie goal hierarchies, drawing into question when and why such pruning is appropriate. When should such processes be dissolved to ensure bodily integrity, or retained through crippling stress in service of progress towards the goal states that they represent, e.g. by attuning to a "new" normal or acting from moral principle, "autonomously" (a capacity in the exercise of which G&dK, 2021a, p. 281, suggest that robots may excel; cf. White, 2020; 2021)?

With change in response to shifting environmental demands associated with Wozniak's "me" and context invariant goals embodied at higher levels of compression of G&dK's bowtie, context-dependent action proceeds via throughput at relatively lower levels, raising the possibility that there might be a sense of self apparent as higher-order throughput potential is not exercised, e.g. "I could do more," or remains yet underdeveloped, e.g. "I can do more," or which most poignantly denies immediate throughput in light of such potential, non-reflectively as an aspect of the embodied situation that is not context dependent, e.g. in the form of conscientious objection, "I will not do that"? A positive answer to this question points to a possible sense of self accompanying each intermediate "me" as one of how context-dependent instances of objective self-determination contribute to or impede actualization of highest levels of a goal hierarchy, aspirations in the realization of which we may associate with so-called "metaphysical" self. This possibility is explored, now.

In metaphysical self, briefly, we are looking for an invariant self-relation across levels of organization from immediate non-reflective to universal moral principle. How might such self-relation manifest in a human being? Some information is available, that the systems in question self-organize in the reduction of uncertainty and with it computational costs associated with tracking unnecessary variables thereby incurring excessive metabolic costs and with this allostatic overload, disease and death. What is necessary is thus a constraint on computation in the service of active inference over the timescales essential to the target architecture, binding personal, intersubjec-

tive, social, cultural-historical and relatively invariant principled moral levels, with such constraint answering to the metaphysical "I" as constant and pervasive, both in a realist neurobiological and in a phenomenal sense of always and already accompanying any given instant of self awareness at more intermediate levels.

Consider in this context Ortega y Gasset's (2002) characterization of self as a constant and pervasive phenomena in terms of "vocation" involving the sense of a globally orienting purpose in life (p. 135). Consistent with the preceding PC inspired review, for Ortega y Gasset (OyG), life is future-oriented, a purposeful self-seeking "program" in pursuit of a target state, "one's life's global project" that also serves as the source of value as objects and others either assist or hinder this pursuit (OyG, 2002, note 149, p. 214). Differently from Metzinger's minimal envelope, experience of one's global project is both essential to and fundamentally directed for OyG, presenting as "pressure" on the "evergoing determination of my present […] exerted on it by my future, i.e. by my vocation or what I have to be, whether I succeed in carrying it out or not (even in part)." (note 158, pp. 215)

Where might such a global project self arise and corresponding phenomenology be grounded in human beings? The default mode network (DMN) stands out as a candidate as it integrates past (memory, hippocampus and related areas) and future (project situations, frontal cortex and related areas) in purposeful imagination of possible situations ("complex goal-directed … memory-based simulations") (Schacter et al., 2012) and in autobiographical memory (Spreng et al., 2009). The "default mode network" was originally so called due to observed suppressed activity during task engagement, with greater suppression during more difficult tasks, and with increased activity in non-action contexts, e.g. mind-wandering. Early research characterized DMN activity as an aspect of shifting action across different contexts, with such activity consistent with recent enactivist accounts of "real-life skilled behavior" in terms of "metastable attunement" as suggested in section 2 of this paper, for example. More recent research has investigated task-related activity in the context of self-appraisal from childhood to adulthood, finding less activation of the anterior DMN especially the dorsal medial prefrontal cortex (dmPFC) during explicit self-appraisal with increasing age corresponding with self-development over the human life course as an aspect of increased functional segregation of anterior (future project) and posterior (actional) components of the DMN, concluding that reduced connectivity correlates with developing self-concept (Davey et al., 2019). One idea here is that implications of instantial self-determinations require less prospection, as expectations are established through prior routine interactions exercised during the life course, as self-concept stabilizes with experience, consistent with the execution of OyG's program as described above and in a process that we may associate with G&dK's (2021a) "maturity."

Adolescent development of the DMN also involves increased segregation from task-positive network activity during a period when cortical potential is highest, decreasing with adult myelination (Park et al., 2021; cf. Vandewouw et al., 2021) i.e. with maturation. Phenomenology characteristic of this development includes accounting for one's self as a social project for an "imaginary audience" in the construction of a personal "fable" (Buis, Thompson, 1989). Narrative self development has been considered the "highest form of cognitive integration" (Hirsh et al., 2013) with "trouble" in the form of challenges to personal convictions a defining aspect thereof (Bruner, 1997). Interestingly, challenges to "protected" values correlate with DMN activity (Kaplan et al., 2017). Recent research distinguishes between two dissociable DMN subsystems, one associated with "valence" and value, and another with "vividness" and detail of prospective (imagined, possible) situations, concluding that the construction of situations (from memory) and their evaluation as worth seeking are neurocognitively separable processes (Lee, Parthasarathi, Kable, 2021; cf. Pezzulo et al., 2021; also, the inchworm and bivalve model of White, 2014). Finally, distinguishable "conservative" and "disruptive" processes modulate development of lasting brain-wide DMN connectivity during adolescence (Vasa et al., 2020). Together, a relatively radical reconfiguration of the whole brain system is experienced including the rapid growth of the prefrontal cortex (and associated mirror systems) responsible for projections over increasingly distant time scales (Fuster, 1989; cf. Pujol et al., 2021). It is worth noting that increased segregation of DMN and task-positive subsystems during adolescence is associated with higher intelligence (Sherman et al., 2014). Indeed, overemphasis on learning engagements with the immediate object environment in education may be undesirable for human childhood development, as this separation may be inhibited (Immordino-Yang et al., 2012).

The proposal here is that the differentiation of developing frontal areas during adolescence from processes embodying action routines and value associations adopted during childhood potentiates the development of a relatively detached, globally orienting project future self-situation. This proposal is complementary to contemporary work in embodied cognition on development of self and consciousness in the context of PC and related approaches. For instance, Anna Ciaunica and colleagues suggest that embodiment within another body during gestation constitutes an "original prior" constraining ongoing development of the organism. On this view, gestation serves to prepare the developing organism for "co-homeostasis" during dependent childhood. And, the present view adds to Ciaunica and colleagues' view that adolescence represents an equally necessary stage wherein individuality emerges in the projection of a uniquely embodied project self-situation (cf. Ciaunica et al., 2021; Ciaunica, Safron, Dellafield-Butt, 2021).

On the present view, life as a global project "I" emerges through more or less normal development of especially the valence associated subsystem of the DMN as a more or less clearly conceived sense of purpose to realize these values in routine interaction with the social and objective world, and around which contextually specific, task-positive subsystems thereafter develop and are organized. Practically, each phenomenal "me" enacted during particular recurrent contexts in life such as when acting as a researcher, a family man, taking care of children, or exercising in a gym, can be represented as a set of sub-attractors of the DMN (corresponding with various bundle accounts surveyed in section 3, above). Valence (answering why these operations are worth performing and refining with increasing precision through directed epistemic agency over the life-course) binding these together develops as a global project "I" that can be characterized as a global attractor with the corresponding sense of self as purpose in life emergent as target valuations segregate from perceptual reality during segregation of developing highest-level default mode from task-positive neural processes.[4] Summarily thus, target state conditions embodied in these processes present as a life-long global project to bring the perceived reality in line with project values, with the felt tension between beginning and end situations accounting for phenomenality answering to Wozniak's "metaphysical 'I' " as well as to OyG's pressure on the present from the future. And, different senses of "me" emerge (including common uses of "I" that Wozniak would classify "me") as each uniquely situated self-seeking program is executed in circular interaction with the shared, objective world, towards embodied project ideals.

OyG's "vocation" answers to Wozniak's metaphysical "I" in the sense of a "lifelong persistent stream of consciousness" as it realizes aspects of itself as component instances of episodic "me" through interaction with a more or less undetermined and under-controlled world. This self-consciousness is not limited to the immediately embodied situation including conformity to social norms, and rather extends across representational time-spans to include invariant values and universal moral principles. Recalling G&C's (2020) relational self-positions within the scope of G&dK's (2021a) evolutionary goals, OyG's greater philosophy emphasizes that each individual occupies a privileged perspective on the shared world with unique potential to contribute to its ongoing determination as a common project through communication of personal experience, making history. The execution of this program, as such, is not a process that is reducible to activities arising between some fractions of brain activity such as might be the case with Wozniak's "delusion" or even necessarily within the confines of an individu-

[4] This is the slope for informative error reduction missing in Miller et al.'s (2021) account of "happiness" as global attractor.

al organism. Rather, OyG's vocational self seems to represent that constant aspect of self brought forward in G&dK's concept of higher-level goals established by an evolved "active inference engine" (G&dK, 2021a, p. 260) amongst other evolved active inference engines with similarly embodied aims.

Interesting in this context, Jesse Bettinger and Timothy Eastman (2017) consider biological cognition in terms of anticipation of self characterized as "predictive model space" that is "counterfactual" in the sense that "the model is an imperfect model trying to optimize its predictions and learn about the system it is modeling" (p. 114). The idea is already a familiar one, that cognition is essentially anticipatory, depending on established neural processes to respond to perceived reality—"information is encoded through synaptic weighting, and the confidence (or precision) of predictions can be altered by hierarchical gain modulation operating as generative models of the system regarding incoming sensory data" (Bettinger, Eastman, 2017, p. 112) — and preparing for most likely outcomes, with "predictions" being "contingent on actual sensory data to become active." (p. 114) Reviewing Alfred North Whitehead and the notion of "proposition," Bettinger and Eastman (2017) distinguish between "prehending" (perceiving) subjects and "logical subjects" in a way reflecting Wozniak's distinction between "me" and "I" mapped onto the perception-action cycle, with "me" upstream and "I" down consistent with preceding discussion (especially of L&F in section 3). On this account, prediction error is fed upstream, becoming the phenomenal "me" while the "I" is characterized as a "might be" on the propositional model of putting forward possibilities (predictions) towards which the living system then pulls itself through action "to maintain a inner-range of state values" evidencing "future-to-present (syntopic, attractor) logic" and apparent "backwards-in-time causality" in contrast with non-living physical system dynamics characterized in terms of "usual past-to-present" efficient causation (Bettinger, Eastman, 2017, p. 117–118), reminiscent of OyG's vocation.

In a way, Bettinger and Eastman capture the intention of the present proposal, with metaphysical self held out as a position to be realized through lifelong self-development. When asking "to what end" such anticipatory systems form, they answer to fulfill "existential needs before those needs become a crisis" (Bettinger, Eastman, 2017, p. 115, quoting Coffman, Mikulecky, 2015) and explore the roles of the salience network in conjunction with midline structures to constrain attention in the exercise of control via allostasis, but limit discussion to retention of individually embodied biophysical integrity. The present view considers the life-course of the organism as modulated by adolescent development as essentially propositional in that a global project ideal is embodied that thereafter serves to constrain cognition to temporally extended values in solution of evolutionary prob-

lems confronting the organism population as a whole, with such ideals informed by development of mirror neural systems thus shaping project ideals in a way that these may embody values independent of individual bodily integrity e.g. principles worth dying for, altruistic aims, rather than local, context-dependent attractors.

Finally, with this comparative account, Wozniak's challenge to "prove that there is a qualitative difference" and "to demarcate the exact border" (Wozniak, 2018, p. 12) between metaphysical subject and its ongoing iterative self-determination can be answered. The image of metaphysical self as essentially propositional places a forward project ideal against more immediate lower-level throughput processes satisfying the stipulation for an invariant self-relation with which this section began. It associates Wozniak's "I" with the feeling of being always and already in the context of progress towards defining values, as a program working to solve what is essentially itself as a uniquely embodied potential solution to evolutionary problems by bringing the perceived reality in-line with project ideals. Ongoing cognition on this model involves testing counterfactuals "as if" actually embodied self-positions (cf. Bettinger, Eastman, 2017; G&C, 2020) in resolution of this embodied project, effectively bridging inherited situations with ideal end states as moderated by adolescent development. Here again, it is important to emphasize the functional segregation of valence and vividness subsystems. With progress towards embodied project ideal, self is objectively determined, and contextually specific "me" related accessible details are embodied with associated processes maturing through iterative interaction toward this end (in this way answering to Wozniak's intuition that PC inspired inquiries into self might focus on so-called "access consciousness" as accessible details are encoded in the context of this metaphysical self-pursuit; cf. Davey et al., 2019).

## 5. DISCUSSION

With OyG's global project, we have the constant and pervasive sense of self answering to Wozniak's "I" that is objectively determined as an embodied inference engine engaging in "what if" processing over self-delimited predictive model space. Why should such higher-level goals develop, outstripping given contextual demands? The idea is that the metaphysical self as project ideal self-situation develops in response to emerging threats to evolutionary goals at levels of organization beyond the individually embodied agent and extending to all similarly embodied (here we may follow Kant in saying "rational") agency not necessarily limited to human agency but deriving from a similar process of adolescent development in other living systems, also (cf. Ledoux, 2021). Accordingly on the present view, self is

essentially purposeful, extending over the course of an anticipated life-span with the potential to represent target situations which an agent may not realistically anticipate inhabiting, e.g. Kant's Kingdom of Ends, though orienting intermediate action towards such ends, regardless.

The notion that the metaphysical self emerges as a globally orienting project binding current with ongoing and future actions in iterative self-determination towards a final project self-position, embodied as a unique proposition that "might be" a solution of evolutionary problems, and that forms during adolescent (highest-level) neural development, allows us to revisit the image of the bowtie. Emphasizing the temporal binding between perceptual instances, we may consider a "traveling bowtie" as one that binds perception and action in a nested goal hierarchy with throughput traveling across levels of processing according to contextual demands and with self actualized in this process. Moreover, consider in this context the heterochronic development of human beings i.e. from prefrontal to hippocampal processes as mediated by the thalamus perhaps specifically the reuniens nucleus (cf. Smaers et al., 2017), alongside changing default and task positive network connectivity, adding another dimension to the journey of the traveling bowtie as it matures over the life-course.

Recalling G&C's (2020) analysis in this light, a "realist" view of something like Wozniak's metaphysical self is constant in a way that is not captured in the traveling bowtie and the context-dependent bundle-theories that it represents. Rather, the "I" per OyG's vocation shapes experience regardless of context. One candidate grounds for constancy in the face of contextual change exists in L&F's (2018) "confidence" in project predictions. Confidence, though phenomenally context-dependent, is constant in that it always involves holding current alongside other, potentially embodied situations, in a way consistent with G&C's self-positional and Bettinger and Eastman's propositional selves, with agents constantly coming to terms with changing situations in a bottom-up and top-down manner (cf. White, 2010, 2014). Another empirical approach which touches on the omnipresence of metaphysical self is available in Andrew and Alexander Fingelkurts and Tarja Kallio-Tamminen's (Fingelkurts, Fingelkurts, Kallio-Tamminen, 2020, 2021) characterization of "witness consciousness" which, like the present view, draws on interplay of subsystems of the DMN. What is absent from witness consciousness is the sense of orientation and with it purpose and source of meaning in life, as with Metzinger's minimal envelope. Absent from L&F (as well as Miller et al., 2021) is how drive to minimize uncertainty informs OyG's vocational call to order apparent disorder from one's unique place in history thereby becoming a solution to G&dK's evolutionary problems through more or less freely directed development of personal potential.

How might such orientation to work, increasing order at personal expense, be best compressed and communicated? Consider the model of

a Platonic cave in complement to the bowtie model architecture, as it articulates a similar input-throughput-output dynamic while capturing the constant orientation to act towards the highest potential of OyG's vocation, making explicit the uncanny sense of self more or less present during routine short-circuits characteristic of metaphysical self according to the present view. There are three sections to the model. The cave represents routine conformity to social norms. The mountain above the cave represents one's highest potential self-situation in the representational space of ideals i.e. Mount Olympus, home of the Gods, corresponding to embodied global project per OyG. And, the reflecting pool on the plane between them under the shade of a tree affords a view of one's self as an object of reflection in front of the mountain behind that "me" representing one's highest calling, which together represent cognizance of purpose in life to achieve that project ideal and satisfy the judgement of the Gods. Ascending and descending the cave corresponds with input-throughput-output on the bowtie model, with the "I" perceived in the difference between the current reflected "me" and who one must become through one's life's work per OyG.

When gazing into the reflecting pool, "I" see "me" recalling Wozniak's resurrection of Wittgenstein's mirror. At the same time, changing focus I can see the mountaintop looming above the surveilled object "me" and against the ideals of which I can feel the space of my progress in self-development towards this highest aspiration. Recalling OyG's pressure from the future on the present, to become what I need to be or fail, one can imagine that it pulls the eyes upward and away from the downward gaze in the direction of the cave which orients agency in norm-seeking and exercising embodied routine per G&dK's short-circuits, effectively toward procedural self-unconsciousness through sufficient precision. G&dK's "personality development" may be seen to involve the balance of forces in either direction, with "maturity" involving the pruning of project ideals cognizant of the homeostatic cave environment in the minimization of stress. On Plato's account, the philosopher who represents this pressure to look upward, communicates the potential above the cave basin and reminds the slaves of their inner duties to seek their true vocations through action, is poorly treated by norm defenders, as the reminder of neglected higher-order goal-states causes stress, as if the cave environment were the one worth seeking, after all (cf. Miller et al., 2021). Here, we may compare Martin Heidegger's "fallenness" to the cave-bound condition, with sense of metaphysical self revealed in the call of conscience that he associates with philosophy (Heidegger, 1998).

The sense of self as outstanding, as propositional, and as represented by the distance between reflected "me" before the mountain-top of one's project ideal, is not captured on the bowtie models. The movement up and down from cave to reflecting pool may be captured by the traveling bowtie. But, the metaphysical self is the view on the present from the summit of

personal potential, OyG's pressure from the future, which on the bowtie model may be represented in the information passed upwards through a goal-hierarchy beyond short-circuit throughput layer, and downwards in comparison of current action against project ideals, delivering OyG's sense of self in the feeling of who "I" must become through a life of directed self-development. This sense of self is captured in the myth of the cave, in the image of the mountain rising above the introspected access of the reflecting pool. Moreover, this image indicates a deficiency in G&dK's assessment of self as relative stability with resilience in terms of maturity, as it allows a focus on the relationship between currently realized or anticipated and highest level project ideals perceived as the pull of OyG's vocation most obvious when one's global project is contravened. The cave model thus affords a focus on metaphysical self as the difference from norms, not realized in norm-seeking entrainment, but in norm-breaking autonomy, instead.

Recalling the situational self of G&C and other multistable routine bundle theories, cave-life represents routine enaction unaware of guiding norms with active and affective mirroring keeping cave inhabitants commonly oriented toward shadows projected by slavers on the cave wall. Given such a shared situation orchestrated for the benefit of others (e.g. politicians, global economic cabals), we can imagine minds "resonating" in coordinated action without prior planning, self-organizing in the common representational space (cf. Pöppel et al., 2021). Yet, there is a reason why this description of life is intuitively unattractive; it is self-nullifying. Far from evidencing a sense of self, the cave model illustrates the loss of self as a standing-out from routine and established norm. It is conceivable then that perceived instability given certain (social) situations is not evidence of mental or personality disorder, at all, and rather that it points to the existence of an outstanding sense of self from which a subject feels a frustrating alienation in the face of especially social pressures to conform to norms that contradict invariant values. This is to say that self as a proposition, in its present situation, is impeded from progressing toward its project conclusion, and the subject may experience debilitating anxiety in the dedication of metabolic potential, trying to compute a way out of the cave if not for one's self then for everyone who stands to suffer for the sub-optimal situation.

In Plato's allegory, a slave may twist at her or his bindings to catch a glimpse of something outside of the play of shadows, to see something of the slavers and their useful idiots who project their propaganda on the cave wall from above. Bound without hope of freedom to seek one's project self-situation, desperation and disorder may result. On this picture, self as project presents in the felt difference between established norms and project ideal with the tension of the chains experienced as stress. This sense of self is directional in that it pulls away from routine expectation, motivating the slave to break from habit, and reclaim one's self from the "they" of

a Heidegger or the "herd" of a Friedrich Nietzsche or the mass psychosis of a Mattias Desmet. This view is also complimentary to Kant's on personality, evident when action towards highest values runs contrary to established routine, and on whose account duty to one's self is experienced as a felt pull upward toward moral perfection, characteristic also of OyG's vocation.

It is in this potential to break free from habit, to stand out from norms of expectation, and to moreover communicate this potential to others who are somehow bound to less, that we may most directly associate the metaphysical "I." So, rather than in seeking resonance with established norms and contemporary expectations, we may identify the feeling of being an "I" in discord, for instance in conscientious objection and the power to say "no" through civil disobedience, extending to construction of popularly accessible accounts in the forms of myths and moral exemplars who die rather than act contrary to highest-order guiding principles, e.g. Martin Luther King Jr., Socrates, Christ. Here, we may offer a word on Jeffrey White and Jun Tani's (2016, 2017) notion of "myth consciousness" originally introduced in the context of cognitive neurorobotics. In that work, "most consciousness" can be associated with Wozniak's phenomenal "me" whereas "myth consciousness" represents awareness of being a metaphysical "I" in ongoing self-development toward ideal situations at most invariant levels of organization, "embodying history in all of its determinations." The cave model represents such a condition.

## 6. RELIGION, AFFORDANCE AND VALUE ALIGNMENT

The traveling bowtie model emphasizes the temporal binding associated with anticipation and prediction, but falls short of shedding light on integrative life-long self-development towards highest-level goal states extending past present personal and social constraints as informed by invariant moral values. This dynamic is captured on Plato's cave model, including also the sense that one has a duty to moral perfection, as in Kant, and the corollary that social norms represent the avoidance of this duty, as in Heidegger's inauthentically "fallen" condition i.e. hiding from one's highest potential behind idle chatter and other distractions characteristic of life in the cave.[5] Self-reflection affords a view on this highest potential, demanding freedom to pursue it through directed self-development—escaping from enslavement to shadows in the cave—as articulated by Plato with his reflecting pool.

[5] Fallenness is natural as routine enaction is not necessarily performed in avoidance of highest duties to self; usually, it is necessary, and a condition which may be associated with G&C's situational self alongside ecological enactivist accounts and other "bundle theories" as surveyed in the present paper. Inauthentic fallenness involves active neglect of highest potentials, including for example composition of academic, e.g. enactivist papers excusing foregone purpose through lack of resolve, the potential for which is not captured on any of the surveyed views.

Clarity on such dynamics affords brief consideration of the practice of prayer and the purpose of religion. Prayer can be viewed as a directed, meditative reinforcement of highest-level goals and iterative increase in precision of ongoing determination of global project-ideal self-situations including inventory of current self-position (in the sense of G&C) relative to project self-position (in the sense of OyG's vocation). This characterization naturalizes prayer as an affirmation of prospects put forward by evolved inference engines as embodied propositions. Prayer on this view can be appreciated as confirming the sense of metaphysical "I" deflated away on purely analytic approaches which fail to capture the intuitive sense of obligation to morally optimal outcomes common to human adolescent self-reports. Again, project self on the present view is a self-organizing solution to evolutionary problems, during and after development felt in the variable commitment to directing personal potential to overcome obstacles to evolutionary goals potentially including highest-possible human situations writ large, i.e. those associated with invariant values and universal moral principles e.g. Kant's Kingdom of Ends as Heaven on Earth. Thus, prayer as a practice of entrainment to evolutionary goals can be seen as prosocial and beneficial to the population of agents across generations not limited to the self-sacrificing Saint or other moral exemplar including potential artificial agents engineered with such capacities in mind (cf. Goekoop, deKleijne, 2021a, p. 281). Ultimately, there is nothing that seems to stand in the way of formalizing such processes, with robot religion providing a computational model proof-of-concept for the importance of faith in human beings (cf. White, 2021).

Here, some note is appropriate regarding moral consideration of artificial agents engineered according to the model of religion as entrainment to prosocial purposes sketched above. Vincent Muller and Michael Cannon (2021) distinguish between context specific "instrumental" and "general" intelligence, and consider that a "superintelligent" general AI may pursue any goal, possibly deviating from human goals, generating the "value alignment problem." Their account proceeds from a decision theoretic characterization of intelligence as a matter of maximizing expected utility, following Stuart Russell (2019). On this view, machines are engineered to optimize performance in specific operational contexts according to reward functions. Concerning current and anticipated technologies engineered accordingly, Muller and Cannon argue that potential value alignment problems derive from human rather than from AI initiatives. Though their distinction between general and instrumental intelligence is interesting as it can be roughly correlated with the functional orthogonality of value and vividness neurosystems considered in this paper, their treatment of "like us" AI reduces to variably broad operational contexts, neglecting value-orienting processes emphasized, herein (consider in this context results of Lee et al., 2021). Instead, their treatment reflects the enactive view from which they build and corre-

sponding characterization of DMN network functionality in terms of multi-stable norm-seeking (as introduced in section 3, above).

The position of the present work is that any intelligence "like us" undergoes different developmental periods (cf. Ciaunica and colleagues' gestation, adolescent development) and that in the process a uniquely embodied ideal goal-state self-organizes in highest-level neural processes that thereafter orients context-dependent engagements according to project values over the life-course. How this project is shaped determines to which values the agent thereafter strives, and what it takes as an opportunity for rewarding action. This view differs from for instance ecological enactivism, as to account for metaphysical self as proposed herein may require a radical revision of that position's Gibsonian take on affordances. Rather than nascent in the environment presenting to clever exploitation, on the present view affordance is better characterized as essentially "self-affordance" because the self as project exposes any genuine opportunity for progress towards its own ideal end as possibly mediated by external-environmental, ecological, factors (cf. Uexkull, 2010). Action motivated otherwise may run contrary to uniquely embodied goals, and so, though an opportunity for (perhaps expedient, norm-satisfying, stress-minimizing) action nascent in the environment, fail to be of meaningful value. To co-opt a popular example, so-called "higher-order" cognition employed to catch a bus to a job in which one is treated disposably by selfish men in the service of short-sighted vision, e.g. money through fraud, is not an opportunity for integrity-preserving action, and rather a chain of enslaving norm. It is not clear thus how ecological enactivists in particular can accommodate the present view without wholesale revision of their position, relocating focus to the internal environment and self-model away from e.g. architectural pre-occupations, relaxing principles for passing pleasure.

Considering "like us" AI in this context invites discussion about freewill in robots and recognition of rights typically afforded human beings on presumption of such potential. Vincent Muller (2021) argues that there is no need to consider robot rights, as contemporary model agents lack freewill and with it a sense of moral responsibility. With no individual locus of responsibility to serve as "bearer of moral status" such agents cannot be afforded rights. Similarly, Keith Farnsworth (2017) argues that freewill requires self-determination understood in terms of organizational closure (being a "Kantian whole") with an internal means for choosing among (more or less available) options according to an agent's "master function," and Rochat (2019) offers a complementary Kantian view that a sense of self-unity as organized and distinct from others is fundamental to any possible learning and experience (cf. Ciaunica, Safron, Dellafield-Butt, 2021). Farnsworth argues that contemporary artificial agents are not "Kantian wholes" in this way, so do not have freewill and with it moral responsibility, thereby

supporting Muller's view on robot rights. For Farnsworth, the master function of biological models is reproduction, ostensibly inconsistent with the present view of metaphysical self which involves pursuing opportunities for action towards an internally self-projected ideal goal-state which may have little to do with biological reproduction. The present view is that Farnsworth's "master function" is directly comparable to OyG's vocation, with the proposal here being that such developmental processes may be formalized for artificial embodiment in the foreseeable future, with robot rights considered accordingly. And, White (2021) argues directly that robots constructed on a Kantian model will be afforded comparable rights when this result is achieved.

## 7. CONCLUSION

The purpose of this paper has been to clarify metaphysical self in the context of contemporary predictive processing (PP) and predictive coding (PC) inspired accounts, to propose possible neural bases for its biological development, to expose how underlying structural dynamics may be represented, and to explore some of the implications of the view for ongoing work in different areas. In the survey of complementary accounts, Metzinger's minimal phenomenal experience (MPE) as the non-"egoic" "natural state" of an agent "predicting itself into existence" (note 26, p. 38, quoting Friston) was challenged. In regards to MPE, Metzinger (2020) concludes "the question of whether and in what sense it can count as 'fundamental,' and whether it is the only truly minimal state of consciousness, has not been answered" (p. 38). The view developed in response is that Metzinger's formalism represents neurosystem dynamics in especially human biological models, providing a kind of analytic envelope, but that the minimal conditions for self experienced as the target sense of a metaphysical subject "I" demands that his MPE envelope be extended in direction of an orienting self-project constraining and directing cognition (cf. Williford et al., 2018). We may consider here the image of a letter composing itself as it delivers itself to its propositional end, with the address on this letter emergent through developmental dynamics during adolescence in human beings. Metzinger's abstract envelope as revealed through meditative practice by developed adults with matured goal hierarchies is fundamental in the sense that it describes (through interoceptive access) the dimensionality of embodied processing associated with enaction, yet it fails to capture that process fundamental to the sense of self corresponding with Wozniak's "metaphysical 'I'" that can be associated with the address on an envelope in transit. In the case of AI, the view here is that such a sense of self and purpose as source of meaning may be formalized in recurrent neural networks constrained by different time-

scales at different levels, with higher order processes modeling increasing invariance associated with context independence and constancy such as in the case of moral principles, and with project aim emerging through developmental processes modeled after those embodied during adolescent development in human beings.

Finally, stepping into the context of current events, it is hoped that this work affords some clarity on the development of sense of purpose and meaning in life for contemporary young people. Recent OECD polling suggests that many adolescents report deficient sense of purpose (OECD, 2019) with nearly half of polled UK adolescents reporting an unsatisfactory sense of meaning in life, for example, raising the issue of the role of education in development of such a sense. At the same time, sense of purpose in life is understood to be protective against developmental disorders and risk taking behaviors including alcohol abuse and sexual promiscuity (Gongora, 2014; Brassai et al., 2011). Given current events (e.g. mandatory masking effectively obscuring crucially mirrored expressions of affect), current interest in understanding human enculturation in order to rectify social injustice for instance in resistance to corrupted leadership (cf. Haslanger, 2019), as well as potential association with personality disorders such as relative instability given changing situations dependent on strength of association with most meaningful, higher-order neural processes that rapidly develop at this stage of biological maturation in life (recalling discussion of G&dK, 2021a, above), policy-makers must be made cognizant of purpose-affirming ends consistent with evolutionary goals. The youth of today are the leaders of tomorrow, tasked with the construction of order in the face of looming disorder at all and increasing levels of organization, from the self on up (cf. G&dK, 2021b). It is ill-advised for policy to run contrary to development of evolved highest potentials, as such would amount to an evolutionary short-circuit and loss of meaning as highest values are contravened, what Michelle Maiese considers "moral atrophy" (Maiese, 2021).

Looking back through history, predictive self-modeling across levels of increasing invariance over increasing time-scales as an aspect of embodied development has reached high-points in the visions of moral exemplars as represented in cultural and mythical heroes and Gods, for example. Pursuit of these ideals has delivered human beings to the present stage. Associations with these high-points and their expressed ideals are lasting, representing invariant values to which persons aspire as propositions and towards which they "predict themselves into existence," thus answering to White and Tani's (2016, 2017) myth consciousness in the felt potential to embody the space of history and all of its determinations. Prayer and some forms of meditation would seem to reinforce these connections, grounding justified resistance to oppression in the call to something greater. The value of such practices to these ends deserves more attention in future work, in the context of artificial

religion and value alignment in artificial general intelligence, and moreover may help to inform current interest of social scientists in ideological oppression with conformity associated with atrophied moral cognition and the correlate dimming of human potential.

## REFERENCES


P. Bagus, Peña-Ramos, J. A., & Sánchez-Bayón, A. *COVID-19 and the Political Economy of Mass Hysteria*, International Journal of Environmental Research and Public Health, 18 (4), 2021; https://doi.org/10.3390/ijerph18041376

J. Bettinger, Eastman, T., *Foundations of Anticipatory Logic in Biology and Physics*, Progress in Biophysics and Molecular Biology, 131, 2017, pp. 108–120.

O. Blanke, Metzinger, T., *Full-body Illusions and Minimal Phenomenal Selfhood*, Trends Cogn. Sci. 13, 2009, pp. 7–13; doi: 10.1016/j.tics.2008.10.003

N. Block, *On a Confusion about a Function of Consciousness*. Behav. Brain Sci. 18, 1995, pp. 227–247. doi: 10.1017/S0140525X00038188

L. Brassai, Piko, B. F., Steger, M. F., *Meaning in Life: Is It a Protective Factor for Adolescents' Psychological Health?*. International Journal of Behavioral Medicine, 18(1), 2011, pp. 44-51.

J. Bruineberg, Seifert, L., Rietveld, E., Kiverstein, J., *Metastable Attunement and Real-life Skilled Behavior*, Synthese, 2021; https://doi.org/10.1007/s11229-021-03355-6

J. Bruner, *A Narrative Model of Self-Construction*, Annal NY Acad Sci, 818, 1997, pp. 145–161.

J. M. Buis, Thompson, D. N. *Imaginary Audience and Personal Fable: a Brief Review*, Adolescence, 24 (96), 1989, pp. 773–781.

A. Ciaunica, Constant, A., Preissl, H., Fotopoulou, K., The First Prior: *From Co-embodiment to Co-homeostasis in Early Life*, Consciousness and Cognition, 91, 2021; https://doi.org/10.1016/j.concog.2021.103117

A. Ciaunica, Safron, A., Delafield-Butt, J., *Back to Square One: the Bodily Roots of Conscious Experiences in early life*. Neuroscience of Consciousness, 2021, 2; https://doi.org/10.1093/nc/niab037

R. W. Clowes, Gärtner, K., *The Pre-reflective Situational Self*. Topoi, 39(3), 2020, pp. 623–637.

J. A. Coffman, Mikulecky, D. C., *Global Insanity Redux*. Cosmos and History, 11(1), 2015, pp. 1–14.

C. G. Davey, Fornito, A.; Pujol, J.; Breakspear, M.; Schmaal, L.; Harrison, B. J., *Neurodevelopmental Correlates of the Emerging Adult Self,* Dev Cogn Neurosci, 36, 2019; https://doi.org/10.1016/j.dcn.2019.100626

G. Deco, Jirsa, V. K., *Ongoing Cortical Activity at Rest: Criticality, Multistability, and Ghost Attractors,* Journal of Neuroscience, 32 (10), 2012, pp. 3366–3375.

S. Dehaene, *Consciousness and the Brain: Deciphering How the Brain Codes Our Thoughts*, Penguin, New York 2014.

K. Farnsworth, C*an a Robot Have Free Will*?. Entropy, 19 (5), 237, 2017; https://doi.org/10.3390/e19050237

A. A. Fingelkurts, Fingelkurts, A. A., Kallio-Tamminen, T., *Selfhood Triumvirate: From Phenomenology to Brain Activity and Back Again*, Consciousness and Cognition, 86, 103031, 2020; https://doi.org/10.1016/j.concog.2020.103031

A. A. Fingelkurts; Fingelkurts, A.A.; Kallio-Tamminen, T. Self, *Me and I in the Repertoire of Spontaneously Occurring Altered States of Selfhood: Eight Neurophenomenological Case Study Reports*, Cognitive Neurodynamics, 2021, pp. 1–28; https://doi.org/10.1007/s11571-021-09719-5

J. M. Fuster, *The Prefrontal Cortex: Anatomy, Physiology, and Neuropsychology of the Frontal Lobe,* Raven Press, New York 1989.

K. Gärtner, Clowes, R. W., *Predictive Processing and Metaphysical Views of the Self*, The Science and Philosophy of Predictive Processing, D. Mendonca, M. Curado, S. Gouveia

(eds.), Bloomsbury: London, UK, 2020, pp. 59–81; https://doi.org/10.5040/9781350099784.ch-004

R. Goekoop; deKleijn, R., *How Higher Goals Are Constructed and Collapse under Stress: a Hierarchical Bayesian Control Systems Perspective*, Neuroscience & Biobehavioral Reviews, 123, 2021a, pp. 257–285.

___ , *Permutation Entropy as a Universal Disorder Criterion: How Disorders at Different Scale Levels Are Manifestations of the Same Underlying Principle*, Entropy 23, 1701, 2021b; https://doi.org/10.3390/ e23121701

V. C. Góngora, *Satisfaction with Life, Well-being, and Meaning in Life as Protective Factors of Eating Disorder Symptoms and Body Dissatisfaction in Adolescents*, Eat Disord., 22 (5), 2014, pp. 435–449; DOI: 10.1080/10640266.2014.931765.

S. Haslanger, *Cognition as a Social Skill*, Australasian Philosophical Review, 3 (1), 2019, pp. 5–25; DOI: 10.1080/24740500.2019.1705229.

M. Heidegger, *Being and Time*, State University of New York Press, Albany, N.Y 2010.

J. B. Hirsh, Mar, R. A., Peterson, J. B., *Personal Narratives as the Highest Level of Cognitive Integration*, Behav. Brain Sci., 36 (3), 2013, pp. 216–217.

J. Hohwy, Michael, J., *Why Should Any Body Have a Self?*, in: The Subject's Matter: Self-Consciousness and the Body, de Vignemont and Alsmith (eds.), MIT Press, 2017, pp. 363–391.

IM. H. mmordino-Yang, Christodoulou, J. A., Singh, V., *Rest Is Not Idleness: Implications of the Brain's Default Mode for Human Development and Education,* Persp. on Psych. Sci., 7 (4), 2012, pp. 352–364.

S. Kahl, Wiese, S., Russwinke, N., Kopp, S., *Towards Autonomous Artificial Agents with an Active Self: Modeling Sense of Control in Situated Action*, Cognitive Systems Research, 72, 2022, pp. 50–62; https://doi.org/10.1016/j.cogsys.2021.11.005

J. T. Kaplan, Gimbel, S. I., Dehghani, M., Immordino-Yang, M. H., Wong, J. D., Tipper, C. M., Damasio, H., Damasio, A., Sagae, K., Gordon, A. S., *Processing Narratives Concerning Protected Values: a Cross-Cultural Investigation of Neural Correlates*, Cerebral Cortex, 27 (2), 2017, pp. 1428–1438.

J. E. LeDoux, *As Soon as There Was Life, There Was Danger: the Deep History of Survival behaviours and the Shallower History of Consciousness*, Phil. Trans. R. Soc. B, 377, 2021; https://doi.org/10.1098/rstb.2021.0292

S. Lee, Yu, L., Q., Lerman, C., Kable, J. W., *Subjective Value, Not a Gridlike Code, Describes Neural Activity in Ventromedial Prefrontal Cortex during Value-based Decision-making*, Neuroimage, 237, 118159, 2021; https://doi.org/10.1016/j.neuroimage.2021.118159

S. Lee, Parthasarathi, T., Kable, J. W., *The Ventral and Dorsal Default Mode Networks Are Dissociably Modulated by the Vividness and Valence of Imagined Events*, J. Neurosci., 41 (24), 2021, pp. 5243–5250.

C. Letheby, Gerrans, P., *Self Unbound: Ego Dissolution in Psychedelic Experience*, Neuroscience of Consciousness, (1), 2017, pp. 1–11.

J. Limanowski, & Friston, K. *'Seeing the Dark': Grounding Phenomenal Transparency and Opacity in Precision Estimation for Active Inference*. Frontiers in Psychology, 9, 643, 2018, pp. 1-9; https://doi.org/10.3389/fpsyg.2018.00643

J. Limanowski, & Friston, K. *Attenuating oneself: An active inference perspective on "selfless" experiences*. Philosophy and the Mind Sciences, 1, 2020, pp. 1–16.

M. Maiese, *Mindshaping, Enactivism, and Ideological Oppression*, Topoi, 2021; https://doi.org/10.1007/s11245-021-09770-1

T. Metzinger, *Phenomenal Transparency and Cognitive Self-reference*, Phenomenology and the Cognitive Sciences 2, 2003, pp. 353–393; https://doi.org/10.1023/B:PHEN.0000007366.42918.eb

___ , *The Ego Tunnel: The Science of the Mind and the Myth of the Self*. Basic Books: New York 2009.

___ , *Why Are Dreams Interesting for Philosophers? The Example of Minimal Phenomenal Selfhood, Plus an Agenda for Future Research,* Frontiers in Psychology, 4, 2013; https://doi.org/10.3389/fpsyg.2013.00746

___ , *Minimal Phenomenal Experience: Meditation, Tonic Alertness, and the Phenomenology of "Pure" Consciousness*, Phil. And the Mind Sci., 1(I), 2020, pp. 1–44; https://doi.org/10.33735/phimisci.2020.I.46.

M. Miller, Kiverstein, J, Rietveld, E., *The Predictive Dynamics of Happiness and Well-Being*, Emotion Review, 14(1), 2022, pp. 15–30; doi:10.1177/17540739211063851

V. C. Muller, *Is It Time for Robot Rights? Moral Status in Artificial Entities*, Ethics and Information Technology, 23, 2021, pp. 579–587; https://doi.org/10.1007/s10676-021-09596-w

A. Newen, *The Embodied Self, the Pattern Theory of Self, and the Predictive Mind*, Frontiers in Psychology, 9, 2270, 2018, pp. 1–14; https://doi.org/10.3389/fpsyg.2018.02270

N. Nyberg, Duvelle, E., Caswell, B., Spiers, H. *Spatial Goal Coding in the Hippocampal Formation*, Neuron, 2022; https://doi.org/10.1016/j.neuron.2021.12.012

OECD, PISA, *2018 Results* (Vol. III)*: What School Life Means for Students' Lives*, PISA, OECD Publishing, Paris, 2019; https://doi.org/10.1787/acd78851-en

J. Ortega y Gasset, *What Is Knowledge?,* State University of New York Press, Albany 2002.

B.-Y. Park, Paquola, C., Bethlehem, R. A. I., Benkarim, O., *Neuroscience in Psychiatry Network (NSPN) Consortium*, Mišic, B., Smallwood, J., Bullmore, E. T., Bernhardt, B. C., *Adolescent Development of Multiscale Cortical Wiring and Functional Connectivity in the Human Connectome*, BioRxiv 2021.08.16.456455, 2021; https://www.biorxiv.org/ content/10.1101/2021.08.16.456455v2

G. Pezzulo, Parr T, Friston K. *The Evolution of Brain Architectures for Predictive Coding and Active Inference*, Phil. Trans. R. Soc. B, 377, 20200531, 2021; https://doi.org/ 10.1098/rstb.2020.0531

J. Pöppel, Kahl, S., Kopp, S. *Resonating Minds—Emergent Collaboration Through Hierarchical Active Inference*, Cognitive Computation, 2021, pp. 1–22; https://doi.org/10.1007/ s12559-021-09960-4

J. Pujol, Blanco-Hinojo, L., Macia, D., Martnez-Vilavella, G., Deus, J., Prez-Sola, V., Cardoner, N., Soriano-Mas, C., Sunyer, J., *Differences between the Child and Adult Brain in the Local Functional Structure of the Cerebral Cortex*, Neuroimage, 237, 2021, 118150; https://doi.org/10.1016/j.neuroimage.2021.118150

P. P. Rochat, *Self-Unity as Ground Zero of Learning and Development*, Frontiers in Psychology, 10, 2019; https://doi.org/10.3389/fpsyg.2019.00414

S. Russell, *Human Compatible: Artificial Intelligence and the Problem of Control*, Viking, 2019.

A. Safron, *The Radically Embodied Conscious Cybernetic Bayesian Brain: from Free Energy to Free Will and Back Again,* Entropy, 23 (6), 2021; https://doi.org/10.3390/e23060783

D. L. Schacter, Addis, D. R., Hassabis, D., Martin, V. C., Spreng, R. N., Szpunar, K. K., *The Future of Memory: Remembering, Imagining, and the Brain*, Neuron, 76 (4), 2012, pp. 677–694.

E. Sennesh, Theriault, J., Brooks, D., van de Meent, J., Feldman Barrett, L., Quigley, K., *Interoception as modeling, allostasis as control,* Biological Psychology, 167, 2022. https://doi.org/10.1016/j.biopsycho.2021.108242.

L. E. Sherman, Rudie, J. D., Pfeifer, J. H., Masten, C. L., McNealy, K., Dapretto, M., *Development of the Default Mode and Central Executive Networks Across Early Adolescence: a Longitudinal Study*. Develop Cogn Neurosci 10, 2014, pp. 148-159.

J. B. Smaers, Gomez-Robles, A., Parks, A. N., & Sherwood, C. C. *Exceptional Evolutionary Expansion of Prefrontal Cortex in Great Apes and Humans*, Current Biology, 27 (5), 2017, pp. 714–720.

R. N. Spreng, Mar, R. A., Kim, A. S. N., *The Common Neural Basis of Autobiographical Memory, Prospection, Navigation, Theory of Mind, and the Default Mode: a Quantitative Meta-Analysis*. J. Cogn. Neurosci., 21 (3), 2009, pp. 489–510.

P. Sterling, Allostasis, *A Model of Predictive Regulation,* Physiology & Behavior, 106 (1), 2012, pp. 5–15.

J. Tani, *Exploring Robotic Minds: Actions, Symbols, and Consciousness as Self-Organizing Dynamic Phenomena*, Oxford University Press: Oxford, UK 2017.

J. Tani, White, J., *From Biological to Synthetic Neurorobotics Approaches to Understanding the Structure Essential to Consciousness*, Part 2, APA Newsletter on Philosophy and Computers, 16 (2), 2017, pp. 29–41.

___ , *Cognitive Neurorobotics and Self in the Shared World, a Focused Review of Ongoing Research*. Adaptive Behavior, 2020; https://doi.org/10.1177/1059712320962158

J. Uexküll, *A Foray into the Worlds of Animals and Humans*, University of Minnesota Press: Minneapolis, MN 2010.

M. Valmisa, *What Is a Situation?*, in: Coming to Terms with Timelessness: Daoist Time in Comparative Perspective, L. Kohn (ed.), Three Pines Press, St. Petersburg, FL 2021, pp. 26–49.

M. M. Vandewouw, Hunt, B. A. E., Ziolkowski, J., Taylor, M. J., *The Developing Relations between Networks of Cortical Myelin and Neurophysiological Connectivity*. Neuroimage. 237, 118142, 2021; https://doi.org/10.1016/j.neuroimage.2021.118142

F. Váša, Romero-Garcia, R., Kitzbichler, M. G., Seidlitz, J., Whitaker, K. J., Vaghi, M. M., Kundu, P., Patel, A. X., Fonagy, P., Dolan, R. J., Jones, P. B., Goodyer, I.M., the NSPN Consortium, Vértes, P.E., Bullmore, E.T. *Conservative and Disruptive Modes of Adolescent Change in Human Brain Functional Connectivity*, Proc. Nat. Acad. Sci. U.S.A., 117 (6), 2020, pp. 3248–3253; DOI: 10.1073/pnas.1906144117.

J. White, *Understanding and Augmenting Human morality: An Introduction to the ACTWith Model of Conscience*, Studies in Computational Intelligence, 314, 2010, pp. 607–621.

___ , *Models of Moral Cognition*. In: Model-Based Reasoning in Science and Technology, L. Magnani (ed.), Springer, Berlin, 2014, pp. 363–391.

___ , *Autonomous Reboot: Aristotle, Autonomy and the Ends of Machine Ethics*, AI & Society, 2020; https://doi.org/10.1007/s00146-020-01039-2

___ , *Autonomous Reboot: Kant, the categorical imperative, and contemporary challenges for machine ethicists*. AI & Society, 2021; https://doi.org/10.1007/s00146-020-01142-4

J. White, Tani, J., *From Biological to Synthetic Neurorobotics Approaches to Understanding the Structure Essential to Consciousness, part 1*, APA Newsl. Philos. Comput. 16 (1), 2016, pp. 13–23.

___ , *From Biological to Synthetic Neurorobotics Approaches to Understanding the Structure Essential to Consciousness, part 3*, APA Newsl. Philos. Comput. 17 (1), 2017, pp. 11–22.

W. Wiese, *Explaining the Enduring Intuition of Substantiality. The Phenomenal Self as an Abstract 'Salience Object'*, Journal of Consciousness Studies, 26 (3–4), 2019, pp. 64–87.

K. Williford, Bennequin, D., Friston, K., Rudrauf, D., *The Projective Consciousness Model and Phenomenal Selfhood*, Frontiers in Psychology, 9, 2018; https://doi.org/10.3389/fpsyg.2018.02571

L. Wittgenstein, *Preliminary Studies for the "Philosophical Investigations", Generally Known as the Blue and Brown Books*. Blackwell, Oxford 1959.

M. Wozniak, *"I" and "me": the Self in the Context of Consciousness*. Front in Psych, 9, 2018, https://www.frontiersin.org/article/10.3389/fpsyg.2018.01656

ABOUT THE AUTHOR — PhD, Philosophy, University Missouri-Columbia, NOVA-LINCS (visiting researcher), Departamento de Informática, FCT/UNL, Quinta da Torre P-2829-516, Caparica, Portugal, and OIST (visiting researcher), cognitive neurorobotics research group, Okinawa, Japan

Email: jeffreywhitephd@gmail.com