New Ideas in Psychology 49 (2018) 27–37

Contents lists available at ScienceDirect

New Ideas in Psychology

journal homepage: www.elsevier.com/locate/newideapsych

Representing is something that we do, not a structure that we "use": Reply to Gładziejewski

H. Oğuz Erdin[a,*], Mark H. Bickhard[b]

[a] Boğaziçi University, Istanbul Turkey
[b] Lehigh University, Pennsylvania United States

1. Introduction

Interactivism has developed into a broad-range framework — from fundamental metaphysical issues to functional brain properties, from neural-glial "societies" to social ontologies, language, and the nature of persons (Bickhard, 2009, 2015a, 2015b; Campbell, 2015). But at its core it is a model of representation and cognition. The interactivist model of representation makes foundational *criticisms* of assumptions concerning representation that have been standard since the pre-Socratics and presents a *positive model* that differs from others on offer in several ways. The interactivist model of representation (or representing), consequently, does not fit well within standard categories (though it is closest to the general pragmatist framework), and, consequently, is often miscategorized and misunderstood.

A recent paper by Gładziejewski (2016) gives us an opportunity to address some of these issues. Gładziejewski presupposes a model of the nature of representation, in a version derived from Ramsey (2007), that is precisely of the sort that the interactivist model argues *does not* work and *is not, in principle, possible*. Gładziejewski argues that the interactivist model does not satisfy the desiderata of this kind of model of representation, not taking into account that the interactivist model argues directly against such models and their desiderata. It is, in fact, a "good thing" that the interactivist model does not satisfy such desiderata, because, if it did, it would be in direct contradiction with interactivism's own arguments that such models are incoherent. The central (though not the only) point is that Gładziejewski begs the question concerning the interactivist model by assuming precisely what the interactivist model argues is impossible, then charging interactivism with not satisfying these impossible desiderata.

It is, of course, possible that the interactivist in-principle arguments against standard models of representation are themselves somehow invalid or unsound, but Gładziejewski offers no arguments against them. By simply accepting the Ramsey framework for understanding and modeling representation, he ignores the arguments against such frameworks, and thereby renders his own arguments circular.

In what follows, we will first give an introductory overview of central interactivist ideas tailored to misconceptions found in Gładziejewski and then track some of the details of his paper, showing how these issues are manifest in it, and attempting to understand how what we take to be errors are motivated by the presupposed (impossible!) framework and desiderata.

2. Central ideas of interactivism

One core principle of the interactivist model is that the metaphysical nature of representation is "being truth-value bearing"; that is, bearing the normative property of being true or false. The central idea is that forms of *functional anticipation of potentialities of agentive interaction* are the loci of such truth values: anticipations of potential interactions can be true or false. It makes intuitive sense that anticipations can be true or false, but much remains to be explicated, modeled, elaborated and defended.

The basic theoretical and philosophical tasks that this core model gives rise to can be parsed into four questions: 1) How does this model account for representation? Is it adequate to the multiform kinds of representation? 2) What are the consequences of the model for other mental phenomena, such as perception, cognition, language, and so on? 3) How are these crucial functions realized in the brain? And 4) What are the theoretical and philosophical challenges that might be encountered by this model, and how can they be met?

These questions have been addressed and elaborated several times. How the model can, for example, account for representation of objects or numbers (Bickhard, 2009). There are, in fact, a number of novel consequences of the model for perceiving, cognition, language, and other phenomena (Bickhard, 2009). The model also yields an account of brain processes that makes sense of phenomena like volume transmitters, non-zero baseline oscillatory frequencies, silent neurons, gap junctions, astrocytes, and so on (Bickhard, 2015a, 2015b). Finally, there is an account of normative emergence, based on an account of metaphysical emergence more generally, which, in turn, is based on an underlying process metaphysics (Bickhard, 2009, 2015c).

There are also multiple challenges that are not of the form "how can the model address X," but, instead, pose critical claims that the model fails in various ways. These could, of course, turn out to be correct, but

* Corresponding author.
E-mail addresses: oguzerdin@gmail.com (H.O. Erdin), mhb0@lehigh.edu (M.H. Bickhard).

https://doi.org/10.1016/j.newideapsych.2018.02.001
Received 4 June 2017; Received in revised form 20 December 2017; Accepted 13 February 2018
Available online 24 February 2018
0732-118X/

we consistently find these challenges to be themselves ill-formed. They often make false assumptions, or, in some cases, assumptions that haved already been addressed and criticized within the overall inteactivist model.

Here we address and correct some such assumptions. The following 8 points should serve as a minimal checklist for any theory or paradigm that comes in close proximity to interactivism.

2.1. Structures can dissolve into system processes

This is a generally unfamiliar idea for someone who is not accustomed to interactivism, but it is "just" a consequence of basic process models, whether machine theoretic or in terms of dynamic spaces. We are so used to thinking in terms of fixed structures and how we use these fixed structures that we have a tendency to structuralize processes in order to render them compatible with our intuitions.

The basic intuition is that data or structure cannot do anything other than influence process, and any such influences can always, in principle, be "built into" the organization of the processes themselves, without any data or structure. Thus structures can always be folded into system processes and disappear as distinct parts of the whole structure, yielding a purely functional theory.

Here, we have one of the most important insight of interactivism: represent*ing* is more fundamental than represent*ations*. Consequently, a strict differentiation between function and representation begs the question. This is a common mistake that pops up in many places, including Gładziejewski's account.

The formalized version of "dissolving structures" comes from abstract machine theory and it is called *the state-splitting theorem*. Consider a state transition diagram with one *register* (i.e. a temporary memory) that can contain a 1 or 0. At some point, the system will *store* a 1 or 0 in the register, and (perhaps) at some later point it will *read* that register and transition in one way if there is a 1 and a different way (perhaps) if there is a 0. Now consider the diagram in-between the store and the read. Split the diagram into two copies, one of which would be entered instead of setting a 1 in the register and the other of which would be entered instead of setting a 0. Similarly, the "1" copy of the diagram will connect with whatever would have been switched *to* if the register had been set with a 1 and the "0" diagram will connect with whatever would have been switched to if the register had been set with a 0. At this point we have a pure transition diagram with no register that is functionally completely equivalent to the diagram with a register. The state-splitting theorem proves that such splitting, though it may become very complex, is always possible, thus, that registers (data, internal or functional structures, etc.) are never necessary.[1]

A version of this is already clear in computationalism: any computer program can be realized in an unbounded range of possible computer architectures; the above point about process is "merely" a generalization of that. In particular, any data structure can be architecturally eliminated by, for example, moving to a pure machine table organization that realizes the (or a) relevant state transition diagram. *Items, memory, data, structure, etc. can do nothing relevant except influence process flow, and those influences can, in principle, always be built directly into the process organization.* The "intuition" to the opposite effect is "just" an error in intuition carried over from thinking about computers (without understanding that they too have unbounded possible architectures that would give you functionally the same computer).

Abstract machine theory, as distinct from computational models, is not here just to provide an example of state splitting. The interactivist account is fundamentally naturalistic and mechanistic in ways that standard computational accounts are not; e.g., the interactivist model does not presuppose representation in its "computations on data" (Bickhard & Richie, 1983). In general, unlike the standard computational approaches, interactivism situates itself in process and control theory, a level of thinking closer to actual machine dynamics than is the case for higher programming principles such as object oriented approaches.[2] We think this is an important advantage both for moving to lower levels (which makes it easier to accommodate lower level mechanisms), hence for naturalizing, and also for unearthing certain key processes underlying standard computational models.

The above point can be made from a *functional* perspective by realizing that data or structure in a system can only influence the functional[3] processing of the system, and that such functional influence does not *require* such data or structure. The overall functional organization can always be rendered as a state transition diagram, or perhaps a machine table (Turing, 1950), and those organizations, in turn, can always be realized in multiple differing architectures. Any given system with data or structures, thus, does not require those data or structures in any semantic sense, nor in any functional sense: the functional consequences of data or structure can always be realized in a different architecture that does not have those data or structures.

This is not to preclude the fact that particular architectures may be useful for purposes of realization: we do not want to use 'simple' universal Turing machines rather than computers. The point, instead, is that any argument that assumes that data or structures are *necessary*, either semantically or functionally, is simply wrong.

2.2. No homomorphism to the external world needed for representations: the rubber band example

Powers (1973) gave a counterexample to our intuitions that there must be some sort of structural similarity with an entity or situation in order to interact properly with that entity or situation: tie two rubber bands together so that they are attached by a knot between them. The experimenter puts a finger through one of the rubber bands, and the subject puts a finger through the other one. The subject's goal is to keep the knot over a spot on the table, adjusting to motions by the experimenter. Clearly, if the experimenter moves too fast, the subject cannot keep up, but, if the motions are slow enough, there is no difficulty maintaining the goal of keeping the knot over the designated 'spot' on the table.

The point comes in noticing that a description for this task will involve concentric circles and spirals around the 'spot', and the elasticity of rubber bands, but that the cybernetic system that is competent to the task will not (necessarily) have any structure whatsoever that is homo- or isomorphic to that task structure. Moreover, this point holds in general for task structures and systems that are competent to those tasks.

Because of such considerations, interactivism does not begin with assumptions of task-competent processes being somehow iso- or homomorphic to task structures, but, instead, begins with general *interaction*-based *differentiations* of the environment. Therefore, interactivist representations, being grounded on (though not simply equivalent to) such differentiation processes, will not necessarily require any iso- or homomorphism with the actual structures in the external world. And that point introduces the next topic.

2.3. Classifications of the environment: implicit definitions, not explicit definitions

One of the key notions in interactivism is *differentiation*. Interactivism distinguishes between *differentiation* and *predication*, and

[1] Bickhard (1980) has an overview of abstract machine theory in Appendix A.

[2] Machine theory, however, does have inadequacies as a modeling language; for example, it can handle temporal sequence, but not timing (Bickhard & Richie, 1983). These can be transcended by moving to dynamic systems theory.

[3] Note that ""functional" is a system-relational concept – an element cannot have a function except relative to something other than itself, relative to some system" (Bickhard & Terveen, 1995, p. 57).

also between *contact* and *content*. In the following two sections we will explain these distinctions.

Consider the following scenario: upon detecting a fly, a frog's internal state will change and stabilize in a determinate end state. Interactivism asks the following question. What other environments produce the same end state? It is well known that frogs flick their tongues to any dark, small, and moving objects. Thus, any such environment, as long as the same end state has been reached, can be classified as a "tongue flicking" kind of environment even though such an environment might contain numerous non-fly objects, such as pebbles or dots on a screen. There is a lot of evolutionary tinkering needed starting from such a coarse differentiation of the environment to a finer-grained differentiation so that it can differentiate, say, between flies and pebbles, or between different types of flies.

The above approach to differentiating environments has the consequence that an end state that has been reached will *implicitly define* the class of environments that could 'cause' it. At this stage we cannot have *explicit* definitions because "there is no semantic information, no representational content, available" (Bickhard & Terveen, 1995, p. 60) to the organism itself although there is a factual correspondence with the class of environments. This is another instance where our intuitions can go awry: in interactivism we have states that constitute differentiating functions without any content! Differentiation is not in itself representation.

Note that, in a homomorphism-requiring paradigm, generalization is the problem which needs to be explained; e.g., how to construct a "chair" representation from encounters with particular chairs. In interactivism, on the other hand, the question is to move from general differentiations to differentiations fine enough to support functionally useful indications of further action and interaction possibilities; so the aim is towards particularization and *not* generalization.[4] This is indeed possible but requires sometimes complex constructions (Bickhard, 1998). Also note that, as a consequence of such coarse beginnings, all the frog can have is a differentiation of the form "a 'tongue flicking' kind of environment." In other words, a classification based solely on frog's internal states and their future possibilities. At this stage only an *external* observer can represent the correspondence between such internal processes and the external world. Although differentiation and predication might *seem* to imply each other, they are crucially not equivalent. This brings us to the next unintuitive property of interactivism.

2.4. Differentiation is not the same as predication: contact vs. content

Interactivism makes a distinction between *differentiation* and *predication*, and argues that differentiation does *not* entail predication: In standard frameworks, if you predicate something (say, being red) to things then you differentiate all those things from other things (that is non-red things) and conversely. Yet in interactivism we have to separate these. What distinguishes predication from differentiation is temporality.[5] In interactivism *differentiation* happens when a certain process reaches its end state as described in the previous item — it differentiates the class of conditions that would support arriving at that end state. This is called *contact* and, as we have said above, there is no content yet. Content comes with predication: "as a predication, a procedure is *indicated as a potentiality*" (Bickhard, 1998, p. 192, emphasis added).[6]

Predications are constituted as indications of possible interactions. The predication is that "this" environment is appropriate for, will support, the indicated interaction. Such anticipations of potentiality have truth value: they can be true or false. The bearing of truth value is the fundamental characteristic of representation.

Such indications also have resources for accounting for more complex forms of representation. One of these is that such anticipatory indications can *branch* into multiple possibilities. The frog, for example, might have several tongue flicking opportunities in several directions — perhaps several flies and a worm. Another resource is that such indications can conditionally iterate: perhaps if the frog were to rotate its body a bit, other tongue flicking opportunities would become accessible: the new opportunities are conditional on the rotation, and the frog might have indications of these linked potentialities.

Such branching and iterating indications can link together to generate vast webs of anticipatory organizations, and do so in complex agents, especially humans. Special kinds of subwebs within such overall webs can constitute representations of objects, and representations of properties realized within those webs can constitute representations of more abstract kinds, such as of the number three (Bickhard, 2009).

Within even the simplest branching organization of anticipations, the organism must select what it will in fact attempt to do. The indications are of what is (indicated to be) available, while the selections within such possibilities will depend on other conditions (e.g., hunger) and goals (e.g., to return to a burrow). Such selection processes constitute the domain of *motivation* (Bickhard, 2000, 2003, 2009).

Note that this means that the indicative relationships cannot be simply causal, nor simply a switching relationship. They have to be indications of possibilities among which motivational selections can take place.

2.5. Content is NOT represented: what are functional presuppositions?

Functional presuppositions are almost trivial when one thinks about an ordinary machine, say, a pen.[7] What conditions should there be in order for it to function properly? Some of them are the following: the temperature of the environment should be between certain limits (e.g., not higher than 1000°), the surface the pen will be used on should have resistance between certain limits (i.e., if it is too slippery the ink will not hold, if it is too high the pen will get stuck), the ink and the chemical composition of its container should not react, there should be enough downward force (e.g. gravity!), and so on. It is clear that although this list begins with trivial considerations, it can easily expand to include so many elements that one can reach all the way up to the whole universe. But a pen nevertheless manages to work quite effortlessly, so we shouldn't worry much about this theoretical issue at this stage.

Now as compared to the standard accounts of representations, there is again an unintuitive turn in interactivism when it comes to functional presuppositions. Assume that a frog sees a fly or a worm. Usually the image thus produced (or some homomorphic structure of it) is taken to be the representation of the fly or worm so that representation then causes a tongue flicking response, or supports an (unconscious) inference to a conclusion that such a tongue flicking is possible. In interactivism, on the other hand, there being a fly or a worm in an appropriate location is the "*presupposed conditions* for the proper functioning of the tongue flicking procedure" (Bickhard, 1998, p. 192, emphasis as in the original).[8] Thus what many theories of

[4] Generalization certainly also occurs, but this is much easier to account for if representation is a matter of differentiation in the first place, rather than, say, impressions of particulars from which universals need to be inferred or constructed.

[5] This is another characteristic of interactivism in that, unlike Turing machine theory, temporality has to be intrinsic (Bickhard & Richie, 1983).

[6] A similar point is that 'concepts' are somehow representations, but they *apply* or not: they do not have truth value per se — "chair" is neither true nor false. An attribution that a concept can be applied, however, can be true or false; can be a proposition, in standard

(*footnote continued*)
terminology. Similarly, an indication that an interaction is possible, based on some differentiation of the environment, can be true or false, but, unlike concepts, the differentiations upon which such indications can be based are not themselves explicitly representational.

[7] With human derived, artifactual, normativity.

[8] Note that anticipating is a functional process, realized in certain kinds of agentive

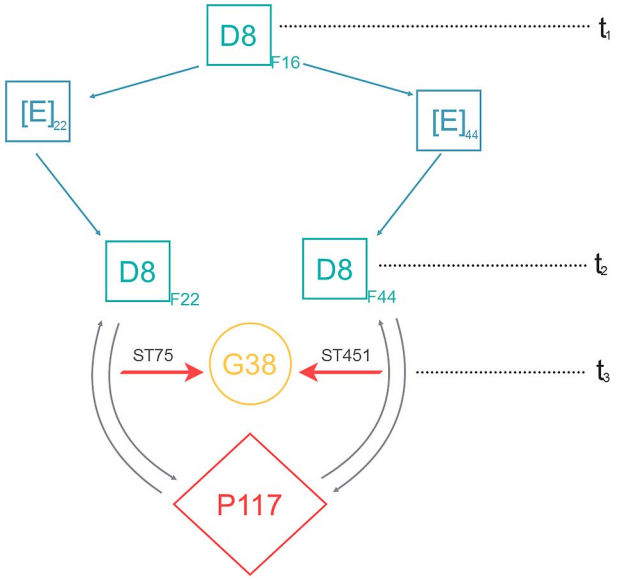

Fig. 1. Differentiation and indication.

representation regard as representational content is "just" a functional presupposition in interactivism.[9]

Dynamic (functional) presuppositions are not "represented" at all in this model — that would be circular — instead, such dynamic presuppositions constitute the *content* of indications that represent that the environment satisfies those presupposed conditions. The representational relationship is with the environment, not with the contents of the representation; to confuse the two is a serious misrendering of the interactivist model and, as we will see, Gładziejewski makes the confusion.

Since the content/vehicle distinction and causal role are the usual go-to notions when discussing various theories of representation, we will make a few remarks regarding their situation in interactivism. In its standard version, the *vehicle* is taken to be a physical structure such as a computer chip's semiconductor gates opening and closing, and the *content* is taken to be what that hardware (or event) is supposed to represent, such as a car or its movement. As we have seen, in interactivism structures dissolve into system processes so that *vehicle as structure* seems to be in jeopardy, and this is indeed the case. If "vehicle" is taken to mean *entity*, then no — that does not suffice.[10] If "vehicle" is taken to mean some sort of functional indication of interactive potentiality, then that's fine. Note that any such function will be realized in some way, e.g., in the brain.

As for the causal role, interactive presuppositions are *modal* (future oriented possibilities) and *counterfactual* (if A holds and B happens, then …) so that they per se do not have a causal role. They are the class of sufficient conditions that would support the interaction if one or more of those sufficient conditions were the case. Presuppositions in this sense can explain why an interaction succeeds or fails, but they do not cause anything (certainly not in a sense of being efficient causes). But the biological processes that realize the functional processes—that is, *anticipations*—certainly do have a "causal" role. Similarly, content also does not have a causal role since presuppositions *constitute* content. Content is also intrinsically modal, counterfactual, and normative. Both presuppositions and content are implicit, not "available" to the system, but success or failure based on those "contents" are available to the system because if the interaction fails, then the presuppositions (whatever they are) are false. This *internal, functional*, detection of interaction failure, thus of presupposition failure, thus of representational falsity, is one of the advantages of being future oriented, not past oriented (Bickhard, 2009).

The normative approach to representations is a crucial point of interactivism and thus a more detailed exposition is presented in the next sub-sections.

2.6. Goals without representing environmental goal conditions: normativity and emerging representing

To sum up: representational predications are indications of potential interaction strategies or procedures. These are predicated of classes of environments which are themselves differentiated via end states of certain other processes — they implicitly predicate that this environment is one of those that would support the indicated interaction, including its indicated potential goals, if any. The indicated interaction procedures have presuppositions about the environment and such presuppositions *constitute* their contents — which are implicit, modal, counterfactual, and, therefore, have no causal role.

With such presuppositions we can see the beginnings of how error (and, crucially, system detectable error) is possible within interactivism. At this point in the discussion, this is error with respect to reaching, or failing to reach, indicated goals.[11] The core point is that indications can be true or false. In more detail, the "basic idea is that other subsystems in the overall system can use the differentiations in those final states to differentiate their own internal goal-directed processing" (Bickhard & Terveen, 1995, p. 61). To give a simple schematized example, consider a differentiator D8 ending up either in final state F22 or else in F44 after interacting with tokens from environment classes E22 or E44, respectively. These end states can be taken by the system organization to indicate the possible interaction strategies St75 or St451 for a process P117 with goal G38. We schematize all this in Fig. 1.

If selection of St75 or St451 fails in achieving G38, then the system can detect that an error has happened somewhere along the line. Such a normative functioning provides the ground for representing to emerge in the possible interaction strategies (St75 or St451).[12]

We have to emphasize that in the figure above there is no label pinpointing the exact location of anticipatory processes. The reason is that "there is a distributivity in interactive representations that is epistemically intrinsic – that could not be altered without destroying the fundamental character of the interactive representations" (Bickhard & Terveen, 1995. p. 305). This is one another aspect of interactivism where one's intutions can be misleading. Thus, a full quote is perhaps in order (Bickhard & Terveen, 1995, p. 305):

> This is a functional distributivity, rather than just an implementational distributivity, involved in the webs of indications of further potential interactions … given an environmental differentiation. Representation is emergent precisely in those functional indications of further potentialities, and those indications are intrinsically distributed within and across the organization of such indications. That is, interactive representation is intrinsically *relational*, and, therefore, necessarily distributed over organizations of such relations.

(*footnote continued*)
(e.g., biological) processes whereas presuppositions are (relational) properties of those (anticipative) functional processes.

[9] The example of the pen is of an artifactual, thus derivative, function, whereas the anticipative function in the interactive model is intrinsically emergent in the nature of the agent. Both involve presuppositions, nevertheless.

[10] Even if it is "just" a bit pattern "entity" in a computer! Even a bit pattern is, for example, a *pointer* only insofar as the (functional) program treats it as such (rather than treating it as, e.g., an integer).

[11] Recall that the fundamental criterion for being representational is possessing truth value — possessing the possibility of being true or false, of the indications being correct or in error.

[12] In this discussion, the normativity is relative to achieving or not achieving a goal. We will see below that there is a more basic form of representational normativity involved in 'just' anticipation per se.

This is in contrast to neural network and Parallel Distributed Processing approaches where the distributivity can be taken as an *implementational distributivity*. It is important to keep in mind this divergence between neural network accounts and interactivism when we begin investigating Gładziejewski's account of Beer's virtual organisms in section 3.2.2.

As a final remark, note that *if the concept of goal requires representations then the whole account will be circular*. The goal subsystems in interactivism can be as simple as a switch (e.g., repeat procedure A with variations if the error is not below a threshold or else switch to procedure B). In other words, goals in interactivism are functional goals of the system and its subsystems, and "the logical function that goals serve … is to provide criteria for error" (Bickhard & Terveen, 1995, p. 63).

On the other hand goals— as 'merely' "final common trajectories" — can be emergent much as in Brooks' subsumption robotics where local interaction rules generate global goals for the robot, such as picking up soda cans in an office (Brooks, 1991). We think that the emergence of goals in Brooks' robots renders the emergence of such final common trajectory outcomes from system processes especially clear. But because Brooks' robots do not have anticipatory processes, they lack interactivist representations properly understood.

It might seem at this stage that interactive representation *requires* goals but this is not the case, and this brings us to the next item.

2.7. Representations without goals: anticipatory functions are all that is needed

In standard approaches to cognition, representation is considered to be an all or nothing phenomenon. In interactivism there is a proliferation of levels. Although we will not present the full hierarchy here (see Bickhard, 1998), it is easy to see that if we get rid of the goal subsystem in the figure above (G38), we can still have *indicative anticipations* with truth values and hence obtain a more primitive form of representation. Actually, the reason we first introduced error within a larger goal subsystem was for ease of understanding.

The key insight is that *truth value does not require goals, only anticipations*. The anticipatory indications may be selected toward the functional service of reaching a goal, and "reaching a goal" is another way in which the overall process can succeed or fail, but such "goal failure" is neither necessary nor sufficient for error per se. It is not necessary because the anticipation could fail, whether or not the goal is reached; e.g., a frog might anticipate jumping into the pond with the goal of escaping a predator without realizing that one of its legs is stuck in a crack. The rotational motion thus resulted from the attempt might nevertheless rescue it from the predator's attack. Conversely, the anticipation may succeed and nevertheless fail to achieve the goal (e.g., a frog in a laboratory might flick its tongue successfully to a dot on a screen without achieving its goal of satiating its hunger). Thus even with goals, it is still the anticipatory function that has presuppositions and truth values.

More formally, normative goal functionality and normative anticipatory functionality must be differentiated, and, thus, it must be noted that truth value does not depend on goal functional success, though it may involve such success (or failure). What is required is normative functionality, and anticipation is such a normative functionality that is more basic than goal functionality. (Meanwhile, "goals" need only be set-points, not representations themselves.)

2.8. Does a soap bubble learn during its adjustments to a wire loop? Neural networks?

Many dynamic systems settle into attractor points or trajectories.[13] At times such attractors have been proposed as constituting representations of any pattern or start condition that is within the relevant attractor basin: that a particular kind of initial condition has been encountered is supposed to be "represented" by the fact that the dynamics have settled into that 'corresponding' attractor.

It should be clear, however, that, while such a dynamic system may *differentiate* among its various possible attractor basins, there is no representation of what is being differentiated. As constituting differentiators, such systems might be useful for a broader system, with the appropriate normativities, for setting up indications of further dynamic possibilities, in which case normative truth value, thus representing, might emerge.

But the differentiators do not themselves represent: what they differentiate is inherent in the dynamics, but the dynamics cannot be "true" or "false." The dynamics settle into a particular attractor or another as a matter of factual contact with the environment. There is no anticipation that could yield emergent truth value (though, again, such an anticipation might be set up conditionally on such differentiations). The settling may be a simple or a complex dynamical process, but it is not in itself normative. If it were, then, for example, a soap film settling into an "equal surface tension" form within a wire loop would constitute representation.

Neural and connectionist nets are dynamic systems that can settle into having certain attractor landscapes via training, which might involve feedback, or might involve "settling" in terms of the statistical properties of the input stream. Although they can be trained, what they ultimately learn is simply correspondences. The nets do not "learn *that* they are correspondences, nor what those correspondences are with" (Bickhard, 1993, p. 295). In this regard, they are no different from standard covariance approaches to representation which require correspondences one way or the other. As we will see, this is also the case for Gładziejewski's two-factor approach. For interactivism, neural networks can be taken as *emergent implicit differentiators*. Because this being only part of what is required, they are *not* learners of emergent representation (Bickhard & Terveen, 1995, p. 301).

If all there is to "learning" is relaxation to an attractor, interactivism will not consider it to be genuine learning. Interactivism requires that the *system learn for itself*. In order for a system to learn for itself, it must have the capability for generating system detectable error — intrinsic normative error, not just 'error' as defined by an external designer or observer — and for that to happen its processes should be of the anticipating nature. It is crucial to note that such anticipations entail a reversal of emphasis from input-to-output dynamic causality to *output-to-input potentialities* (Bickhard & Terveen, 1995, p. 294). The system must anticipate its output and the interaction of that output with other surrounding processes should be the new input. In other words, the flow must be inherently circular, not sequential. In this sense representation "is fundamentally a matter of anticipation, and not a matter of a system being a retrospective spectator back down its input sequence" (Bickhard & Terveen, 1995, p. 304). Note that approaching the issue of representation from the general issue of "how learning for a system itself is possible" again led us to render representations in terms of indications of potential interactions, as we were arguing was the case in the previous items.

3. A critique of Gładziejewski's charge against interactivism

3.1. An outline of Gładziejewski's argument

Gładziejewski claims that representations as depicted in action oriented theories are too unconstrained because they do not meet Ramsey's *job description challenge* (Ramsey, 2007; we will abbreviate the challenge as JDC). Such accounts are too liberal in admitting certain structures or systems to be representational which should not be considered as representational according to other norms. Assuming this argument works, Gładziejewski goes on to offer a remedy, namely a

[13] Which can, in certain circumstances, be strange attractors.

two-factor theory of representation: something is a representation if it guides actions by exploiting its structural similarity to whatever it is representing (Gładziejewski, 2016, p. 23; subsequent page references are to this article, unless otherwise noted).

Gładziejewski's aim is not to determine "whether ACToRs [action guidance theories of representation] give us a good theory of how intentional content is determined" but rather to "concentrate on their *explanatory* value for cognitive science" (p. 17, emphasis as in the original). He draws on Ramsey's JDC: In a given theory of cognitive science, look for the functional roles of the purported representations; this is their *job description*. Then, evaluate whether this job description "can be classified—in an intuitive, natural and understandable manner—as playing the role of representation" (Gładziejewski, p. 17). Ramsey's own examples of what does *not* meet his job description challenge are receptors and Dretske's (1988) notion of representation. Here we have to be careful about the second step in JDC because the critique of any candidate depends on the criterial norms being applied. Gładziejewski does not give a clear enough account of them at the beginning of his paper, but at this stage his criteria include:

1. It should make "sense to describe these structures as *standing-in* for something" (p. 17, emphasis as in the original). The entities worthy of the name "representation" should function as the entities "we pretheoretically recognize as representations—like maps, fuel gauges, or sentences of natural language—function" (p. 18). In other words, they should be like *external* representations.
2. It should be indispensable; it should "enable us to understand phenomena in way that is impossible without it" (p. 17).

This second criterion will be the key reason for accusing ACToRs of being too liberal.

In section 5 where he outlines his two-factor theory, Gładziejewski is clear that structural similarity is what he has in mind (see especially his footnote 8 on page 23 for a clarification). He aims for a necessary and sufficient condition for something to be a representation: "action guidance is by itself insufficient for making something a representation, what other conditions need to be met?" (p. 22). Certain conclusions and certain steps in his arguments during the earlier parts of his paper can only be understood when one realizes that he has this agenda in mind.

Now before moving on, we would like to anticipate a little about the general form of our critique. Notice a fork in his approach: if by "functional role" he means the functional role of "representation" as a *concept* used within cognitive science and philosophy of cognitive science (as a metatheoretical consideration), then the above intuitive criteria seem to us totally ad hoc. We will return to this issue in section 3. If, on the other hand, he means representation's functional role as a *structure* within a theory, then the first item in our list of interactivist ideas (structure dissolving into process) comes in conflict with his account presented so far. According to his description of JDC, one has to highlight the purported representations and *then* look for their functions. But this assumes a differentiation between function and representation. Consequently he is already rejecting a central feature of interactivism; namely, that structures can dissolve into processes, and that representing emerges in a kind of functioning. He is begging the question against the interactivist model, whose claim is that representing is itself an emergent kind of functioning.

Gładziejewski agrees with the criticisms of purely correspondence-centric theories of representation as given by Bickhard (1993; Bickhard & Terveen, 1995). On the other hand, the reason he finds purely action-centric theories lacking is that they fail "to give us an idea about what distinguishes representational action-guidance from nonrepresentational action-guidance" (p. 23). We will respond to this criticism later. First, we should delve into his criticism of interactivism proper. We are not addressing his evaluation of Rosenberg and Anderson's action guidance theory of representation (Anderson & Rosenberg, 2004).

3.2. Gładziejewski's critique of interactivism

Gładziejewski presents two arguments for showing that interactivism does not meet the JDC for its concept of representation. The first one is a conceptual argument and the second one is empirical (Gładziejewski, 2016, p. 18). We will begin with the conceptual argument.

3.2.1. Gładziejewski's conceptual charge

Gładziejewski's main idea is that the interactivist definition of representation, as indications of interactive potentialities, is a strongly reductive one, which attempts to "explain representational normativity with the normativity of action" (p. 18). This reduction in turn renders representation talk explanatorily useless because everything "said using representational talk could be more economically said by reference to actual or possible (inter)actions and their conditions of success" (p. 18). One should take extreme caution here because one of the central points of interactivism is emergence. Hence, there is no straightforward way to replace representational talk with action talk in such a way that the former becomes explanatorily useless.

Gładziejewski is moving too fast here. One might as well say that since atoms are composed of elementary particles like quarks, it is explanatorily useless to keep "atoms" in chemistry. If the issue is merely pragmatic, there is great use of atomistic talk, even talk of Newtonian physics where it is appropriate. If Gładziejewski intends an in-principle argument, then substantial issues are either not addressed at all or only insufficiently handled. In Gładziejewski's view, Bickhard should be so ruthless a reductionist that his theory renders all talk of representations superfluous. In fact, the opposite is true. Representations (indications of interactive potentialities) are needed (for any complex agent), and they emerge whenever the internal control processes reach a certain kind of complexity such as when more than one option for interaction becomes possible or when the interfaces between subsystems need to cope up with the required complexity. When it comes to learning and memory, even stand-in types of representations are needed (Bickhard, 1998).[14] It is quite ironic that a theory emphasizing emergence almost everywhere is accused of strong reductionism.

Gładziejewski's argument for reduction seems to be along the following lines:

> According to interactivism, those indications are internal structures or processes that (1) are activated or arise before the organism engages in particular action (say, predator avoidance); (2) can, and sometime will, lead to the organism actually performing this action; (3) have semantic or representational contents that are determined by the conditions of success (dynamic presuppositions) of the action to which they can lead. (p. 18)

According to him (3) follows from (1) and (2) because "the representational status … is fully constituted by the role they play with respect to guiding or preselecting action" (p. 18). But here we have to ask: the representational status of *what*? This is another instance of the assumption that representation is structural, so the issue (for Gładziejewski) is how those structures are functionally *used*. There seems to be no recognition of the *emergence* of representing (and not representation) in certain kinds of functioning for interactivism.

Although these three properties do apply, they are highly insufficient and potentially misleading from the interactivist perspective. Note that properties (1) and (2) are true for computational symbol processing models too. Symbols are activated and processed, culminating in a final decision before they are sent to outputs which then lead the computer or the robot to actually perform an action. The first two properties are near-universal for any model that posits structure for

[14] But note that they are not quite "standard" representations because the contents are not explicit.

some level of cognition. Only the third seems specific to interactivism.

But then how can one infer that "indications have property (3) in virtue of having properties (1) and (2)" (p. 18)? At best (3) follows from (1) and (2) *if the intrinsic normativity of action is recognized and taken into account*, but Gładziejewski doesn't do that; instead he treats action as a simple motion or cause, and then smuggles normativity in as "success." If action is not taken as normative, then (3) does not follow from (1) and (2). There have to be other criteria that are missing in his account, which in turn lets him accuse interactivism of reductionism. We have presented some of these criteria in section 1 above.

Perhaps more fatally, the problem is that property (3) is stated so vaguely that it might be acceptable per se! For example, if "determined by" is taken to mean "caused by" or "inferable from," then the clause seriously misrepresents the interactivist model. If "determined by" is understood as "constituted by," then it might be correct. Gładziejewski ignores how dynamic functional presuppositions are *intrinsic* to interactive anticipations, and so misconstrues the interactivist account of content. He is missing its emergent core.

Now, focusing on the third property, Gładziejewski says that the "simple fact that some internal activity could eventually cause an action whose success depends on environmental conditions gives, by itself, no leverage to the idea that this activity *represents* those conditions" (p. 18, emphasis in the original). But for interactivism the relevant activities do not in any way *represent* the conditions they functionally presuppose. Rather, those 'activities' emergently *bear* those assumptions *about the environment*. So, the representational relationship is with the environment, not with the content about that environment. Unfortunately, as we mentioned in the previous section, he does not give clear enough criteria at this stage for what makes something represent something else. Thus, we have to go back to our list above.

If Gładziejewski is taking the first item in the list, then at least a necessary condition for him is that the representation be a stand-in. But this makes the charge vacuous because, other than Ramsey's job description challenge, he does not present any arguments as to why one should stick to such a norm for representations. Then, the question becomes as to why JDC requires pre-theoretical and intuitive notions of representation such as having structurally similar inner models, especially in the light of all the problems that beset such standard notions. For this reason it seems that he is begging the question. As mentioned above, Gładziejewski himself agrees with those charges. But the "charges" that Gładziejewski agrees to are in-principle, and thus his move is to a model that (he agrees, apparently) is in-principle impossible. That is, Gładziejewski seems to agree with the critique that correspondence, thus structural, models of mental representation cannot be correct, but then advocates precisely such a correspondence/structural model.

Finally, the accusation of offering an over-liberal account of representation follows as a consequence of the first argument. Once one accepts that everything has been reduced to causal-action talk by a species of ACToR, one might think that there are many types of non-normative action guidance.[15] Being non-normative, such action guidance would "certainly" not be representational. So, if action guidance that is not representational seems possible on this construal of non-normative action, one ends up being too liberal in attributing representations to such "actions." But if the first argument is unsuccessful, as we have claimed, this consequence falls with it too.

To sum up, Gładziejewski's argument at this stage seems to be merely that a representation must be a stand-in because of JDC, and interactivism's representations are not stand-ins. He then goes on to offer two more arguments (again short ones) against a possible response to his accusation that interactivism is reductive hence that there is "no reason for claiming that interactivism gives us an explanatorily valuable notion of representation" (p. 18). He begins by stating that "interactive representations are *not* explicit, but rather represent their contents only *implicitly*" (p. 18). He then presents an argument from Ramsey that implicit representation does not entail inner, structural representation, and, thus, does not satisfy the JDC.

First of all, note that, as mentioned more than once up to now, the contents are *not* represented. So, the locution "represent their contents" indicates a confusion on Gładziejewski's part. It is not the representing of contents that constitutes something as representational: it is the possessing of (presupposed) contents about something else (e.g., the environment) that makes something representational.

Gładziejewski mentions that presuppositional content is implicit, then references Ramsey's arguments concerning implicit representation, and then concludes that "this does not entail that there is anything even remotely functionally resembling a representation *inside* the system" (Gładziejewski, 2016, p. 19, emphasis in the original) — that is, there is no entailment that there is an isomorphic structure inside the system. It is not clear that Ramsey's notion of "implicit" is the same as the interactivist notion,[16] but, setting that aside, it is certainly correct that interactivist representation does not require isomorphic structure inside the system. Citing that as a criticism of the interactivist model, however, thoroughly begs the question whether structural isomorphism should be criterial for representation at all. The interactivist model holds that the most basic criterion is that of bearing truth value, not of being interpretable as an isomorphic correspondence. So, we welcome the conclusion that implicit presuppositional truth value does not entail structural correspondence. The reason Gładziejewski has not found anything remotely resembling (structural) representations is because he is looking for something in a place where it is by definition not there.

His second argument takes the form of a *reductio*. Even if one accepts implicit representations (Of the stand-in type? There also seems to be a confusion in that interactivist content is implicit, but representation in the sense of indications is *not* implicit), one still "needs to provide a good rationale for using representations talk at all" (p. 19). He then reconstructs a possible interactivist response as follows:

1. Each indicated action has dynamic presuppositions.
2. Those presuppositions constitute the content of representations in interactivism.
3. "There is nothing in the system that would explicitly represent this content. (p. 19). The representations are indications of potential interactions that "enable the organism to perform actions" and this "is the *only* way in which those indications are related to external conditions" (p. 19).
4. Hence, all the content must be implicit in interactivism.

For this argument to work as a *reductio,* implicit content must be unacceptable. But implicit content is unacceptable, in this argument, only if structural isomorphism is criterial for representation. As mentioned several times so far, we do not accept that criterion, so relying on it begs the question.

Furthermore, we can accept that this is a partially correct characterization of the interactivist position — content is most fundamentally constituted in implicit presupposition. Nevertheless, interactivism does allow for the existence of explicit representations. *Second level representations* that interact with the first level do render explicit properties that are only implicit in the level below. This requires further model construction within the theory and perhaps that's why it is missing in Gładziejewski's article. Because of time and space constraints we cannot provide further details here but the relevant constructions are available in Bickhard (1998).

[15] It's not clear what non-normative action guidance would be, unless the "action" at issue is itself non-normative, and, thus, not really action. For us a mere, say Newtonian, motion without normativity does not fall under the category of "action."

[16] There are, in fact, several notions of "implicitness" within the interactivist model per se (Bickhard, 1998).

He continues:

1. Adding to this mechanism that "they *also represent contents*, even implicitly, simply adds nothing genuinely new or explanatorily valuable to the picture" (p. 19).
2. This "merely presupposes representations" (p. 19) so that the question is still being begged.

Once again, Gładziejewski assumes that representations represent their content. The interactivist will deny this. For that matter, it should be denied by any candidate model of representation: if representation is supposed to be constituted in representing content, no matter how this is supposed to be the case, the model will be circular — representation (of content) is presupposed in (supposedly) modeling representation. Thus, since content is not necessarily represented in interactivism, the above charges are the ones that are indeed begging the question. Moreover, given that the conclusion in 2 is the complaint made by the interactivists against standard approaches, we were surprised that Gładziejewski makes the same charge against the interactivists. We think that this is only possible because he assumes that the notions of representation in the two different paradigms can be compared and interchanged in a straightforward manner. For him representation requires structural similarity, whereas for us such a property (when it does exist) can be derived from more basic processes, so that it is not a necessity but a possibility; a possibility which could have evolutionary or functional advantages, in some circumstances.

Before we move on to Gładziejewski's empirical charge, one final remark about the nature of structural similarities in interactivism. The sense in which the intuition of homomorphism — structural correspondence, or stand-in — can be most accurately realized within interactivism is a *modal* sense: the organization of the *possibilities* of interaction given a system for engaging in such interaction, for successful (true) presuppositions (thus representation), is homomorphic to the organization of *potentialities* afforded by the environment. Thus modal trajectories creatable by the organism can correctly "trace" modal possibilities afforded by the organism-environment relationship. That is, there is a kind of modal homomorphism involved. But, as mentioned in item 1.2 of the first section, there never needs to be a functionally explicit structure that is homomorphic to any actual structure in the world.

3.2.2. *Gładziejewski's empirical charge*

For his second charge against interactivism, Gładziejewski presents Randall Beer's (2003) anti-representational virtual organisms as meeting interactivist criteria for having representations, thus rendering interactivism too liberal to meet the JDC (Gładziejewski, 2016, p. 14). We will first delineate why he thinks Beer's organisms have no representations. Then we will consider why he thinks interactivism must endow Beer's organisms with them.

Beer's virtual organisms have a three level neural network which lets them eventually classify two different kinds of objects and engage in appropriate 'actions': avoid diamond-shaped objects and catch circles (hybrids of these objects are also involved, but they are not important for our purposes here). Gładziejewski agrees with Beer in declaring these agents non-representational; so, they conclude, any representational talk in this setting is explanatorily useless. One reason for this is that "the virtual agent does not make use of any kind of model of its environment" (p. 19). Also, there are "no internal, causally relevant structures governing its actions that correspond to those categories" and "there is nothing inside it that would meet the JDC" (p. 19). Thus, we don't need to postulate representations to explain anything about these creatures. According to Gładziejewski representation necessarily requires a correspondence-based inner model. If there is no such inner structure, there is nothing representational. This rules out, by assumption, any other approach to representation, and in particular the interactivist variety of representation. So the question is being begged, independent of whether interactivism per se is right or wrong.[17]

Must interactivism acknowledge representation in Beer's organisms because they satisfy the interactivist criteria for having representations? Gładziejewski has two criteria for his claim:

1. The ability to select from one of two possible actions.
2. The existence of internal states and processes at the middle layer that precede action. These processes decide whether the organism should catch or avoid the falling object. They can be thought of as indications of potential interactions. Gładziejewski cites Beer: "his results […] suggest that the 'decision' [about how to act] is repeatedly made and unmade as the agent and the object interact until the organism eventually "commits" and actually performs one of the actions" (p. 19).

To these we should perhaps add the following just to make his presuppositions clear:

3. This is a "goal" directed system.
4. It interacts with its virtual environment and has "learning" capabilities (the reason for the scare-quotes in these two items is that from an interactivist point of view these organisms have neither genuine goals nor learning capacities, as we will see below).

Because of these Gładziejewski gives the following argument whose conclusion is that interactivism should grant representations to Beer's organisms (p. 19):

It seems, then, that interactivism would have us think that in this case there are patterns of internal activity that (1) precede action and (2) could potentially lead to the agent performing a particular action; thus we should say that these patterns indicate interactive potentialities that therefore (3) *represent* the action's conditions of success. It follows that Beer's virtual agent is a non-representational system that is categorized as representational by Bickhard's theory.

Would these suffice for a minimal representational system according to interactivism? The answer is no. There are several reasons for why they fail the interactivist criteria.

First, Gładziejewski yet again has representations representing their contents.

Second, Beer's robots lack the normativity required for genuine representations: their internal states are *differentiations* but not *anticipations*. What Gładziejewski has been describing of Beer's account holds for any dynamical system that relaxes into one of its attractors (see item 1.8 in section 1). If this is the case, then there is no representation in Beer's organisms in the interactivist sense. If interactivism granted anticipations to any dynamical system relaxing to one of its attractors, then Gładziejewski would be right in his criticism of it being too liberal. In fact, there have been many interactivist criticisms of connectionist and other network models many times for exactly this reason (e.g., Bickhard & Terveen, 1995).

In such neural networks mere input processing cannot detect genuine error for the system itself because any input is just like any other input. There cannot be any normative distinction between non-error sensory input and pure error input — the inputs influence dynamics in differing ways, but there is no anticipation. Here interactions are not "being epistemically essential to the *constitution* of the representations" (Bickhard & Terveen, 1995, p. 293). In other words, there is no genuine normativity here and hence no anticipatory processes as Gładziejewski concludes in (3) in the above quote.

The adjusting to 'error' — the 'learning' involved — is an abuse of

[17] Reminder: the interactivist model argues that correspondence models are incoherent as fundamental models of representation (though there may be derivative forms of representation that can be understood in correspondence terms), and *Gładziejewski seems to accept those arguments.*

language—there is 'error' only in the interpretation of the observer or designer. To claim that the adjustments in, say, a back-propagation system constitute genuine learning would require also saying that the dynamic adjustments of a soap film across a wire loop constitute 'learning' (see item 1.8 in the first section). To repeat, the issue, then, is "what constitutes learning *for a system itself*."

Perhaps a general lure of the standard neural network architectures and their learning capacities is the following simplified analogy. In nature, brains resemble neural networks and the designer norms provided to these networks might be claimed to be given to the brains via innateness or by some early developmental constraints provided by the environment. Since learning is genuine in nature, this analogy implies that neural network learning could be taken as a genuine learning also. But as we have seen, such networks lack anticipations and thus normativity. According to interactivism the lack of normativity in such standard architectures is an indication for a need of a radical shift in design principles.[18] Furthermore, we find that the analogy to brain processes is not entirely accurate. Because of time and space constraints the reader is referred to Bickhard (2015a, b) for a model of brain processes which is compatible with interactivism.

Third, according to interactivist criteria, Beer's system is at most a basic *differentiator* of two environments. As mentioned in items 1.4 and 1.8 in section 1, this is not enough for representations although it is a required first step. Indeed, this is a general critique of all claims that attribute representational states to neural networks from the point of view of interactivism (Bickhard & Terveen, 1995). Although there are unsupervised nets also, such nets still relax into states depending on the statistics of the input flow, and there is no normativity. Moreover, they have to be designed so that the statistics that they pick up on are going to drive the relaxation dynamics in the designer-desired manner. So, for any neural network to be a candidate for having representations, they not only have to be more than input classifiers according to a designer-specified norm, but their interactions with the environment must also be inherently normative and necessary to the constitution of representations.

Fourth, focusing on Beer's case now, Gładziejewski's discussion seems to us to be violating item 1.1 in section 1: it presupposes that representations have to be structures or singular elements. The interactivist model does not require that indications are realized in structures, so Beer's models could be accepted in terms of their not having singular structures. That is, Beer's model could be accepted as a simulation of one aspect of the interactivist model — a kind of "minimal" simulation of *differentiation* — but it cannot be a realization of the model per se. So, it is not the case that "Beer's virtual agent is a non-representational system that is categorized as representational by Bickhard's theory" (p. 19). Beer and Gładziejewski both assume that representations have to be singular items of some sort; that is simply wrong and rejected insofar as representing is an anticipatory activity and is modeled as such. So, again, the charge against the interactivist model is question begging.

In general, although we find Beer's models interesting, questions as to whether they realize represent*ing* are multiply confused and confusing. They do not contain correspondence structures, but that is OK for interactivism. They also do not contain any indicative structures or elements, but that is also irrelevant: indication or anticipation is functional, not structural, and need not be realized in singular elements or structures at all. So Beer's models do not violate anything in the interactivist model on this point either. Although they do not *violate* anything in the interactivist model, it should be clear by now that they are *missing* many crucial aspects that they would need to have representations or genuine learning.

[18] E.g., metanets where the connections of a given neural network are themselves nodes of another network so that the connections can be terminated or added or otherwise modified by the metanet itself might be a step in the right direction (though far from sufficient). This makes it possible to change the whole graph topology of the network in addition to mere adjusting of weights. Note that metanets are a more accurate model of real brains where new connections come and go depending on the learning involved (see Bickhard & Terveen, 1995, pp. 301–307 and the references therein). Here too, however, there is still no normativity.

4. A critique of Gładziejewski's two-factor theory

After presenting his critique of ACToRs, Gładziejewski offers a two-factor theory of representation. He claims that since action guidance by itself is not sufficient, one needs action guidance that uses representations which have structural similarities to whatever they are representing (Gładziejewski, 2016, p. 25). Although, by now, it should be fairly clear how an interactivist critique will proceed against any theory with such a requirement of homomorphism, we will nevertheless proceed with it in the hopes of further illustrating certain aspects of interactivism.

One motivation for augmenting action guidance theories is the following (p. 22):

> "What is missing from this picture is a good idea — or any idea — about the nature of *differentia specifica* that distinguishes representational action guidance from the kind of action guidance that is achieved without employing representations."

This illustrates a danger of using the notion of "action guidance" instead of "interaction indication" — 'guidance' does not carry the normative anticipatory meanings that "indication" does, so, unless guidance is further explicated in terms of anticipatory indications, it can easily look like there is nothing normative, thus nothing representational, involved.

Gładziejewski gives an example where the task is to navigate from point A to point B in a city and he presents three different methods of accomplishing this task (p. 22). The first way is just by asking a local to guide you through the streets. The second way is by following a trail from A to B, such as by following small red balls. The third way is by using a map where A and B are marked on it. Gładziejewski claims that only in this last instance can the required differentiation between representational and nonrepresentational action guidance be established (p. 22):

> I propose that the difference-making factor here is the fact that in order to succeed, person 3 exploits what we might generally call a "correspondence" between the map and the terrain. More precisely, what I think makes the case of person 3 a case of representation-use is the fact that (1) this person (a representation user) uses a map (a representational vehicle) to guide her action with respect to the terrain (what is represented), and (2) she does it by employing a strategy whose navigational (action-guiding) success is non-accidentally dependent on whether a certain type of relation holds or holds to a sufficient degreed between the map (representational vehicle) and terrain (what is represented).

It is unfortunate here that the example is explicitly one of using an external representation. Such external representations do exist, and they do provide a source for the 'resemblance' relation between representation and represented in the person making use of the representation (as well as the source for normativity in that person). But it is precisely the representing *that is being done by that person* that we are attempting to model. Inner representing cannot be of the same form as external representation: external requires interpretation while internal cannot, on pain of unbounded regress. This issue resurfaces at multiple points in the article when Gładziejewski makes such claims as "the lesson is that representations are things that succeed in playing their action guiding function by exploiting a certain relation between the representation itself (the vehicle) and what it represents" (p. 23). Our immediate reaction is to ask, exploited by whom? What is needed is for these neural (and other) activities to modulate other activities in *anticipatory* ways.

It is even more unfortunate for Gładziejewski to claim that a "map is a useful action-guider only if its structure matches or resembles — or to the degree it matches or resembles — the structure of that which it represents" (p. 23). There are at least three problems here:

1. Such "matching or resembling" can be determined only from the perspective of an external observer (unless it is rendered in terms of successful interaction indication/guidance, in which case it is either false qua structure or it amounts to the interactivist model).
2. Everything matches/resembles everything else if the 'appropriate' 'mappings' are used — that is, what counts as structure is not determinate without a readout or interpretive process, and that is what is at issue in the first place.
3. The only functional consequence that 'structure' can have is to influence the flow of process, and any such influence can always be incorporated into the organization of the system dynamics without any such structure at all. In other words, this aspect of his claim is against item 1.1 in section 1. Perhaps even more fatally, it is against item 1.3 too, the rubber band experiment: a goal-directed cybernetic feedback system can adequately "guide" task performance when the task has a particular structure that has no natural relationship with the organization of the cybernetic system at all.

Recall that the central point of the interactivist model is that interaction anticipation is what constitutes the *emergence* of representational normativity from functional (or pragmatic) normativity. If Gładziejewski recognized this central aspect of the interactivist account, he would, presumably, address how it is possible to have "practical error" regarding interaction anticipation without having truth value, and, therefore, without having representation. Instead, he seems to subsume the interactivist model under the "guidance" model with no normativity, and then claims that guidance per se is not enough — you need *structural guidance*. But (and again) the only support for this is his repeated invocation of the JDC, and, at least with regard to the interactivist model, that completely begs the question. Even on its own terms, this is a strikingly unsupported, ad hoc, criterion. At best, it makes an appeal to our intuitions as to what we would want to call a(n *external*) representation (no 'represent*ing*' as primary here); at worst, it involves deep confusions about how structure can be defined, and how structure can influence process. Our question at this level is: What if that "job description challenge" is itself based on false presuppositions? This would seem to be the core point.

One final remark before concluding this paper is that Gładziejewski mentions Grush's emulation theory of representation (Grush, 2004) as "a genuinely representational explanation that uses the notion of internal representations as mechanical or automated 'models' or 'maps'" (p.24). Bickhard has published a critique explicitly directed at Grush's model (Stojanov & Bickhard, 2004). Moreover, as mentioned before there is no problem with structures constituting representations so long as there is some normative dynamics to interpret those structures; this is as true for internal 'structures' as it is for external 'structures'. Also, Grush's model does not actually satisfy the JDC because the "structures" have no (necessary) structural homomorphism with what they are used to 'guide' — they are anticipatory (though there is no account of normativity), they are useful for 'fast' anticipations, and such anticipation does not require structural correspondence.

5. Conclusion

Interactivism, as a model of representation, has many novel aspects that are inconsistent with the strongly ingrained intuitions of standard and long standing theories of representation. However these counter-intuitive aspects of interactivism are needed to overcome the insufficiencies and incoherencies of the available models of representation. We believe that one fundamental lesson of the history of physics is that our most basic intuitions can be transformed or even abandoned if the available theories turn out to be inadequate. For these reasons, and in the hopes of making our responses to Gładziejewski more accessible, we have presented some of the key aspects of interactivism in the first section.

In the second section we have outlined Gładziejewski's argument, and his conceptual and empirical critiques against interactivism. There are four main issues that we are finding problematic. First, missing the emergent nature of representations in interactivism, he accuses it of being reductive to action only. According to him interactivists must then grant representations too liberally to too many systems, obliterating the explanatory value of theories of representation. As it is hopefully clear from our response, the emergent core of interactivism is necessary for its ambitions. Representations are not structures but emergent processes out of interactive processes, hence the title of this article: Representing is something we do, not a structure that we "use."

This brings us to the second issue in his account. Gładziejewski is committed to the necessity of a homo- or isomorphic idea of representation as a stand-in. This in turn corners him into thinking that representations have to be structures, which means that they are "used" somehow. We not only find such a structural similarity requirement for representations ad hoc, but for many reasons insufficient, and, ultimately, incoherent. In order to overcome such insufficiencies, interactivism has had to draw distinctions such as contact versus content or differentiation versus predication. We have delineated these in the first section and some additional explications are scattered through the article. As we have outlined in section three, Gładziejewski finds an interactivist account insufficient and "enhances" it to a two-factor theory where there needs to be both structural similarity and action guidance that exploits this in a suitable manner. Besides the infinite regress that always arises with models based on analogies with external representations—just as an external representation needs a cognitive agent to interpret it, the cognitive agent's internal representation will need a homunculus, an internal agent within the agent, to interpret it, then the homunculus' internal representation will need a second homunculus, an agent within an agent within an agent, to interpret *it*, and so on ad infinitum—and the ad hoc nature of his account, we presented three further problems with such an approach.

Third, there seems to be a neglect on Gładziejewski's part regarding the importance of normativity and its naturalization in interactivism. This comes to the fore in his empirical charge against interactivism (second subsection of section two) via Beer's virtual organisms. According to interactivism, representations have to have a normative aspect and the only way to naturalize normativity without obliterating it is via anticipations. We pointed out that if anticipations are reduced to a dynamical system's usual progression towards one of its attractors (and the same applies to neural networks), then one cannot speak of normativity. We argued that since, according to interactivism, Beer's neural networks are at most differentiators (albeit ones that "learn" to differentiate) with no anticipatory processes, then Gładziejewski's empirical charge using this kind of neural network begs the question against interactivism.

Fourth, key properties and relations of *content* from an interactivist standpoint seem to be missing in Gładziejewski's account. In both sections two and three we have tried to emphasize that dynamic functional presuppositions are *intrinsic* to interactive anticipations and the activities of the organism do not *represent* those conditions and, thus, content. The fact that the representational relation is not with the content but with the environment seems to be a general misunderstanding that has surfaced in multiple places.

We conclude that the interactivist model is *not* vulnerable to the criticisms posed by Gładziejewski, while his two-factor model *is* vulnerable to criticisms of correspondence models. We hope that our explications as to why that is so will promote better understanding of the interactivist model.

Acknowledgements

We would like to thank Lucas Thorpe for reading and commenting on our drafts. This project is partially supported by the Tübitak 1001 project titled “Concepts and Beliefs: From Perception to Action” with grant number 114K348.

References

Anderson, M. L., & Rosenberg, G. (2004). A brief introduction to action guidance theory of representation. *Proceedings of the 26th Annual Conference of the Cognitive Science Society,* 1185–1190.
Beer, R. D. (2003). The dynamics of active categorical perception in an evolved agent. *Adaptive Behavior, 11*, 209–243.
Bickhard, M. H. (1980). *Cognition, convention, and communication.* New York: Praeger.
Bickhard, M. H. (1993). Representational content in humans and machines. *Journal of Experimental & Theoretical Artificial Intelligence, 5*, 285–333.
Bickhard, M. H. (1998). Levels of representationality. *Journal of Experimental & Theoretical Artificial Intelligence, 10*, 179–215.
Bickhard, M. H. (2000). Motivation and emotion: An interactive process model. In R. D. Ellis, & N. Newton (Eds.). *The caldron of consciousness* (pp. 161–178). Amsterdam: J. Benjamins.
Bickhard, M. H. (2003). An integration of motivation and cognition. In L. Smith, C. Rogers, & P. Tomlinson (Eds.). *Development and motivation: Joint perspectives* (pp. 41–56). Leicester: British Psychological Society.
Bickhard, M. H. (2009). The interactivist model. *Synthese, 166*(3), 547–591.
Bickhard, M. H. (2015a). Toward a model of functional brain processes I: Central Nervous System functional micro-architecture. *Axiomathes, 25*(3), 217–238.
Bickhard, M. H. (2015b). Toward a model of functional brain processes II: Central Nervous System functional macro-architecture. *Axiomathes, 25*(4), 377–407.
Bickhard, M. H. (2015c). The metaphysics of emergence. *Kairos, 12*, 7–25.
Bickhard, M. H., & Richie, D. M. (1983). *On the nature of representation: A case study of James Gibson's theory of perception.* New York: Praeger.
Bickhard, M. H., & Terveen, L. (1995). *Foundational issues in artificial intelligence and cognitive science: Impasse and solution.* Amsterdam: North-Holland.
Brooks, R. A. (1991). Intelligence without representation. *Artificial Intelligence, 47*, 139–159.
Campbell, R. J. (2015). *The metaphysics of emergence.* New York: Palgrave Macmillan.
Dretske, F. I. (1988). *Explaining behavior.* Cambridge, MA: MIT Press.
Grush, R. (2004). The emulation theory of representation: Motor control, imagery and perception. *Behavioral and Brain Sciences, 27*, 377–442.
Gładziejewski, P. (2016). Action guidance is not enough, representations need correspondence too: A plea for a two-factor theory of representation. *New Ideas in Psychology, 40*, 13–25.
Powers, W. T. (1973). *Behavior: The control of perception.* Chicago: Aldine.
Ramsey, W. (2007). *Representation reconsidered.* Cambridge: Cambridge University Press.
Stojanov, G., & Bickhard, M. H. (2004). Representation: Emulation and anticipation. *Behavioral and Brain Sciences, 27*(3), 418.
Turing, A. M. (1950). Computing machinery and intelligence. *Mind, 59*, 433–460.