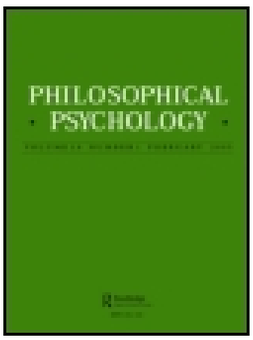

Philosophical Psychology

ISSN: 0951-5089 (Print) 1465-394X (Online) Journal homepage: http://www.tandfonline.com/loi/cphp20

Rethinking the explanatory power of dynamical models in cognitive science

Dingmar van Eck

To cite this article: Dingmar van Eck (2018): Rethinking the explanatory power of dynamical models in cognitive science, Philosophical Psychology, DOI: 10.1080/09515089.2018.1480755

To link to this article: https://doi.org/10.1080/09515089.2018.1480755

Published online: 22 Jun 2018.

PHILOSOPHICAL PSYCHOLOGY
https://doi.org/10.1080/09515089.2018.1480755

Routledge
Taylor & Francis Group

ARTICLE

Rethinking the explanatory power of dynamical models in cognitive science

Dingmar van Eck

Department of Philosophy and Moral Science, Ghent University, Ghent, Belgium

ABSTRACT

In this paper I offer an interventionist perspective on the explanatory structure and explanatory power of (some) dynamical models in cognitive science: I argue that some "pure" dynamical models – ones that do not refer to mechanisms at all – in cognitive science are "contextualized causal models" and that this explanatory structure gives such models genuine explanatory power. I contrast this view with several other perspectives on the explanatory power of "pure" dynamical models. One of the main results is that dynamical models need not refer to underlying mechanisms in order to be explanatory. I defend and illustrate this position in terms of dynamical models of the A-not-B error in developmental psychology as elaborated by Thelen and colleagues, and dynamical models of unintentional interpersonal coordination developed by Richardson and colleagues.

ARTICLE HISTORY
Received 22 June 2017
Accepted 26 February 2018

KEYWORDS
Causal explanation; cognitive science; dynamical model; explanatory power; mechanism

1. Introduction

Perspectives diverge on the explanatory structure of dynamical models in cognitive science (and other scientific domains). Some favor covering-law interpretations of dynamical models (Bechtel, 1998; Walmsley, 2008), others argue that some (though not all) models rather are mechanistic in character (Kaplan & Bechtel, 2011; Zednik, 2011), a different set of authors claims that they are causal explanations (Gervais & Weber, 2011; Weber, Van Bouwel, & De Vreese, 2013), and yet others – in the context of (systems) neuroscience – offer minimal model interpretations of dynamical models (Chirimuuta, 2017; Ross, 2015; Silberstein & Chemero, 2013; Woodward, 2017).

A related point of dispute concerns the explanatory power of dynamical models in cognitive science (and other scientific domains). For some, dynamical models have explanatory power to the extent – and only to the extent – that they refer to underlying mechanisms. If they fail to do so, they may have predictive and descriptive value, yet not explanatory force (Kaplan & Bechtel, 2011). Others take it that predictive power, under

CONTACT Dingmar van Eck Dingmar.vanEck@UGent.be Department of Philosophy, Ghent University, Blandijnberg 2, Gent 9000, Belgium

certain circumstances, is sufficient for explanatory power (Stepp, Chemero, & Turvey, 2011; Van Gelder, 1998),[1] or that a measure of explanatory power is secured by the fact that dynamical models convey scientific understanding, namely that qualitative consequences (insights into how a system behaves or will behave in a certain way, i.e., what it does or will do) can be derived from these models (Gervais, 2015).

A key issue concerns whether or not dynamical models in cognitive science have explanatory power when they do not refer to underlying mechanisms. I argue that most extant defenses of the explanatory power of "pure" dynamical models in cognitive science – ones that do not refer to underlying mechanisms – fail to secure the explanatory value of such models. Prediction is not sufficient for explanation, and neither is bona fide understanding secured when dynamical models are completely unconstrained by mechanistic evidence, that is, when there is no evidence at all that there are underlying mechanisms that operate in accordance with the processes described in the dynamical models. I offer a novel, broadly interventionist, account of the explanatory power of pure dynamical models in cognitive science.[2]

I argue that some pure dynamical models in cognitive science are *contextualized causal models* and that this explanatory structure gives such models genuine explanatory power. A main element of the model explanation concerns a causal claim that cites a core causal factor that makes a difference to the explanandum. Another important element concerns (mathematical) descriptions of (internal and external) constraints and their relations with the core causal factor, which specify when the core causal factor is (and when it isn't) a difference maker for the explanandum phenomenon. These (mathematically described) *contextual dependencies* between constraints and core causal factor (s) are counterfactual dependencies – answer what-if-things-had-been-different questions – and thus show what the value of the explanandum phenomenon depends upon (in a broadly Woodwardian sense): changing the values of the constraints (captured by mathematical parameters) in these dependencies leads to changes in the values of the explanandum phenomenon (also captured by a mathematical parameter). This counterfactual element gives such explanations genuine explanatory power (despite the fact that they do not specify mechanisms).

I take constraints, in general, to refer to task conditions and differentiate between internal and external constraints. Whereas external constraints refer to specific characteristics or properties of the task environment under which a task is (to be) performed by subjects, internal constraints refer to specific restrictions on the behaviors that subjects (are to) execute during a task. These constraints are represented as parameters (and not as variables) in the dynamical models considered in this paper.

I elaborate this view in terms of dynamical models of the A-not-B error in developmental psychology (Smith & Thelen, 2003; Thelen, Schöner, Scheier, & Smith, 2001) and dynamical models of unintentional interpersonal coordination (Richardson, Marsh, Isenhower, Goodman, & Schmidt, 2007).[3]

Let me stress that the above models are "pure," that is, non-mechanistic, ones. There is, I argue, a principled difference between these dynamical models and mechanistic models (be they "complete," schemata, or sketches). These dynamical models list dependencies between external constraints and causal factors. Since such external constraints are not, on any reading of constitutive relevance, constituents of mechanisms, they are not articulated in mechanistic models. *Ipso facto*, dependencies between external constraints and core causal factors are not specified in mechanistic models. Therefore, dynamical/contextualized causal models are importantly different from mechanistic models.

Furthermore, mechanistic information, specifically localization and manipulation information with respect to mechanisms, is not required to *explain* the explananda phenomena *and their contextual variability*, targeted by contextualized causal models. Rather, contextual dependencies do the explanatory work in explaining when a core causal factor is or isn't a difference maker for a target explanandum. However, I argue that such mechanistic information is relevant from an evidential point of view. Evidence that the operations specified in dynamical/contextualized causal models actually (or probably) are component processes of mechanisms or constituted by mechanisms confers plausibility on the specification of operations in these models and, relatedly, information on mechanism manipulations is relevant to back up the claim that the core causal factors listed in contextualized causal models indeed are difference makers for the targeted explananda phenomena.

The view that I elaborate in this paper thus agrees with those extant perspectives that stress the relevance of mechanistic commitments, but conceives of this relevance in different fashion: mechanistic commitments are evidentially, rather than explanatorily, relevant.

The structure of the paper is as follows. In Section 2, I briefly outline dynamical explanations and discuss the current controversy concerning the explanatory status of dynamical models in cognitive science. In Section 3, I elaborate my alternative account, according to which some dynamical models are "contextualized causal models." I illustrate the explanatory insights that such models provide in terms of dynamical models of the A-not-B error in developmental psychology. In Sections 3 and 4, I contrast my view with extant perspectives on the explanatory status of pure dynamical models in cognitive science. A key result is that these models can be explanatory, despite the fact that they do not describe mechanisms. In Section 5, I discuss a second case, unintentional interpersonal coordination, in order to show that the analysis applies more broadly to dynamical models in cognitive science. Section 6

contrasts the account advanced in this paper with a related but distinct project, the regimentation of dynamical models as non-causal minimal models in the context of systems neuroscience, and offers conclusions.

2. Dynamical models in cognitive science

2.1. *Dynamical explanation*

Dynamical systems theory is a conceptual framework for studying dynamical systems, in which concepts that capture key features of such systems are related and elaborated in dynamical models of such systems in terms of mathematical formalisms, typically difference and differential equations. Dynamical systems theory has been applied to study phenomena in diverse scientific fields, such as physics, economics, neuroscience, and cognitive science. The term 'dynamical system' is used in a variety of scientific and mathematical contexts and, consequently, there are many different interpretations of the term 'dynamical system' (Van Gelder & Port, 1995). Here I use the (slightly adapted) characterization given by Van Gelder and Port (1995), which they deem most useful for characterizing *natural cognitive systems* and the dynamical approach toward studying such systems. On their characterization, dynamical systems are taken to be sets of changing properties (they use the term 'aspects') of the world in which changes in properties are induced by changes in one or more of the other properties. Some properties thus interact with one another in this change-inducing sense (a specific change in a specific property can result in several concurrent changes in other properties. These other properties then merely change at the same time, as a result of the change in that specific property, without there being a change-inducing relation between these other properties). Furthermore, systems at any point in time exhibit a state, which results from the values that the properties belonging to the system have at that point in time. Depending on the values of the properties, there are a large number of states the system can be in. The totality of these possible states makes up the state space of the system. The behavior displayed by a given system is interpreted as the state changes the system goes through in real time, that is, its trajectory through the state space of possible states (Van Gelder & Port, 1995). State changes can result from property (variable) changes, that is, changes of property (variable) values, within the system, and changes in the system can result from factors outside the system that affect or constrain the system's trajectory through the state space. These external factors are called 'parameters'. Parameters affect or constrain change in the system, but not vice versa (Van Gelder, 1998; Van Gelder & Port, 1995).

Properties of dynamical systems are captured in dynamical models of those systems as variables (with certain values) and these variables and

their values are related in terms of mathematical equations. These equations describe how the behavior of a system evolves in real time as a result of the way in which changes in variable values impact values of other variables, and they may capture how parameters affect or constrain change in the system. These equations are considered (by most) to be law-like mathematical regularities.

According to the "dynamical hypothesis" in cognitive science, cognitive systems are dynamical systems (Van Gelder, 1998). On the dynamical view, cognition is understood as a continuous and dynamic interaction or coupling between organism and environment unfolding in real time. Cognitive scientists endorsing this view, *dynamicists*, for short, apply concepts from dynamical systems theory to the study of cognitive systems. They aim to describe, explain, and predict the behavior of cognitive systems, that is, the state changes the system goes through in real time, in terms of dynamical models of these systems, specifically, in terms of the equations governing the relations between changes in variable values in these models.

Phenomena that have been extensively studied from a dynamical systems perspective include, among others, coordination dynamics phenomena like rhythmic finger movements (Haken, Kelso, & Bunz, 1985), circadian systems phenomena like circadian oscillations and anticipatory synchronization (Stepp et al., 2011), and developmental psychological phenomena like the A-not-B error in 7–12 months old infants in developmental psychology (Thelen et al., 2001). This latter phenomenon has been the subject of numerous investigations, in dynamical terms, of the research group of Thelen, Smith, and collaborators.

The A-not-B error concerns the phenomenon, seen in 7–12 months old infants, where infants continue to reach to a location where they have uncovered a toy even when they see in subsequent trials the toy being hidden at a new location (Thelen et al., 2001). Piaget (1954) was the first to describe and attempt to explain the phenomenon, which he did in terms of developmental stages with respect to the acquisition of knowledge of object permanence. According to Piaget, only from the age of 12 months onwards do children possess the knowledge that objects can exist independently of the actions they direct against or perform with them. Following his influential work, several other, sometimes competing, explanations for this phenomenon have been advanced over the last 60 years or so. These alternative explanations were developed because Piaget's view fails to accommodate the finding that the error, although very robust in the canonical experimental format, is highly sensitive to contextual effects that disrupt the occurrence of the error (e.g., [slight] changes in the visual properties of the hiding location). If the error were a faithful measure of the manner in which objects are represented by 7–12 months old infants, that is, without object permanency

being a property of the representation, how can it be that such infants perform like 12 month olds, that is, do not commit the error and thus act as if object permanence is a property of their object representations, when contextual conditions are (slightly) altered? In Piaget's wake, alternative (developmental) explanations have been formulated to account for the error and its contextual subtleties, focusing on, among others, shifting representations of space when infants grow older, age-related changes in functionality of prefrontal cortices, changes as regards response inhibition, and the strengthening of representations when infants grow older (Smith & Thelen, 2003; Thelen et al., 2001).

Commenting on these explanations, Thelen and colleagues state:

> Each of these accounts captures some truths about infant perseverative reaching, but none has a full explanation of both the canonical error, *and* the richly documented effects of context which are part and parcel of the same phenomenon. (2001, p. 4).

Here is the most important dissimilarity to other explanations of the A-not-B error and the key to the dynamical explanation advanced by Thelen and colleagues (2001):

> We deeply disagree with the widely held assumptions that knowing and acting are modular and dissociable. Indeed the cornerstone of our dynamical model is that "knowing" *is* perceiving, moving, and remembering as they evolve over time, and that the error can be understood simply and completely in terms of these coupled processes. (p. 4).

Motor memory is at the heart of the dynamical model explanation:

> Infants make perseverative location errors because the motor memory of one reach persists and influences subsequent reaches. (Thelen et al., 2001, p. 9).

So motor memory of previous reaches is, in this explanation, the key causal factor that makes a difference to the occurrence of the error. This factor is represented in a dynamical model as mutually interdependent or coupled with processes of looking, motoric planning, and reaching. These factors/processes and their couplings and the manner in which they unfold in real time are formalized in terms of differential and difference equations and represented as trajectories through a state space. More specifically, Thelen and colleagues (2001), based on extensive experimental findings, simulated the decisions of infants to reach to the A or B location by means of a model, which they call the 'dynamic field model' (p. 10), in which the various factors influencing those decisions are related and their impact on one another captured by means of equations.

The key equation of the model reads as follows (Thelen et al., 2001, p. 17):

$$\tau \dot{u}(x,t) = -u(x,t) + S(x,t) + g_{intra-field}[u(x');x']$$

X stands for the movement parameter, $u(x)$ for its dynamic field, which represents the values of the reaching movements of the infant during the A-not-B task in terms of points on the field. Each point on the field corresponds to a particular spatial location of the baby's arm during the A-not-B task. When the activation level of a specific point surpasses a particular threshold level, a reach is induced toward the corresponding (A or B) location. This dynamic field changes continuously with time *t*, captured by $u\ (x, t)$. Task inputs (structure of the task, the reaching cue [A or B] and, after the first reach, motor memory of previous reaches) are represented by $S(x, t)$ (more on this below). *T* refers to the time scale parameter. *T*, together with $- u(x, t)$, characterizes the time scale over which the field builds up or decays. $G_{intra\text{-}field}$, finally, captures cooperativity, that is, interactions between points on the field that either enhance or inhibit activation levels of the field. Interactions may initiate from any point *x'* on the field. *G* corresponds (psychologically) to an age-related developmental parameter that affects the ability of infants to execute accurate reaching movements (perseverative reaching is, *inter alia*, age dependent: in the canonical situation, it is observed in infants 7–12 months of age. Older infants do not display perseverative reaching in the canonical situation). This model thus articulates, in terms of the above equation, how the activation level of every point (x) on a dynamic motor planning field (u) changes over time (t) as a function of the field's previous activation (_u), an input vector (S), a cooperativity parameter (g), and a temporal decay constant (T plus $- u(x,t)$).

The task input driving the continuous evolution of the field comprises three sources: specifications of the task environment (e.g., properties of the hiding locations of the toy, such as ambiguous vs. salient, distinct locations), the cue to either reach to A or B, and (values of) motor memory which, after the first reach, biases subsequent reaches. The action decision that evolves in the motor planning field ($S\ [x, t]$), based on contributions from these three sources, is formally captured by means of the following equation (Thelen et al., 2001, p. 18):

$$S\ (x, t)\ =\ \mathrm{S}_{task}(x, t)\ +\ \mathrm{S}_{specific}(x, t)\ +\ \mathrm{S}_{memory}(x, t)$$

S stands for the motor planning field, *x* is the movement parameter, and *t* captures the "foundational assumption" (p. 16) that the field continually changes with time. $S_{task}\ (x,\ t)$ is the parameter representing the task environment, $S_{specific}\ (x,\ t)$ the parameter representing the cueing event, and $S_{memory}\ (x,t)$ the parameter capturing the contribution of the motor memory of the system on the current motor decision $S\ (x,\ t)$.

Motor memory is a key factor: "Most crucial for this account is that once infants reach, a memory of that reach becomes another input to the next trial" (Smith & Thelen, 2003, p. 345; see Thelen et al. 2001, pp. 8–9).

The stronger the memory becomes (due to repeated cues over time to a certain location), the stronger the influence of motor memory on the decision to reach to a certain location. (So, for instance, an infant cued [correctly] to location A on three trials is more likely to make the error on the fourth trial, i.e., continue to reach to A even though the target is now hidden at location B, than an infant who is cued fewer times to location A.)

The values that can be assigned to the parameters and variables in the dynamic field model are based on extensive experimental findings. So for any given experimental situation, the model offers an explanation of the error (and its variations) in terms of the values of the parameters and variables which specify contextual dependencies or constraints in which motor memory of previous reaches (the core causal factor) is or isn't a difference maker for the A-not-B error. The model expresses the fact that perseverative reaching is highly dependent on such constraints as the relative ambiguity of the task input; if the containers, that is, the hiding locations A and B, are visually distinct, infants are less likely to commit the error. The same goes with respect to the delay between looking and reaching; as mentioned, motor memory of previous reaches does not cause the error when there is no or very little delay between looking and reaching. Likewise, posture changes affect the likelihood that motor memory of previous reaches causes the error; 8–10 months old infants are less likely to commit the error if they go through the trials standing rather than sitting (more precisely, when they are being held in an upright, standing position by adults).

A key feature of the model is thus that it accounts for the error, as well as contextual variations influencing its occurrence (or absence):

> Because all the processes contributing to the behavior are coupled, continuous, and based in time, we can account in one model for both the error itself and for the decline in perseverative reaching *in different situations and at different ages*. (Thelen et al., 2001, p. 4).

Phrased in (broadly) Woodwardian (2003) terms, the model answers what-if-things-had-been-different questions, that is, the model makes clear how the error, as well as its decline depends on (values of) contextual parameters – on contextual dependencies between the core causal factor and internal and external constraints.[4] These dependencies are captured by means of a number of equations, which represent how changes in the values of parameters affect the value of the explanandum phenomenon, the presence or absence of the A-not-B error. For instance, how changes in the delay between looking and reaching affect whether motor memory of previous reaches is a difference maker for perseverative reaching, or not. I further elaborate on this issue in Section 3 where I address the explanatory structure and power of such models. First I discuss a number of other perspectives on the explanatory structure and power of (pure) dynamical models.

2.2. *Diverging perspectives on explanatory structure and power*

The use of dynamical models is widespread in cognitive science. There is, however, ongoing controversy about the *explanatory* power of dynamical models in this domain (as in neighboring fields like neuroscience). As Kaplan and Bechtel (2011) rightly assert, "the real pivot point for the debate centers on what makes explanatory dynamical models explanatory" (p. 439). Is the model developed by Thelen and colleagues (2001) genuinely explanatory? Answers diverge depending on the perspective one adopts. Importantly, although Thelen and colleagues (2001) take their model to be biologically plausible, it is a pure dynamical model that does not reference mechanisms and yet is considered explanatory nonetheless. By their lights, their model is explanatory despite – or perhaps because of – the fact that "The model as it stands is neutral as to an anatomical instantiation in the central nervous system: is it a model of the behavioral dynamics" (Thelen et al., 2001, p. 28).

Whereas, Stepp and colleagues (2011) recently argued, by appealing to prediction, unification, and counterfactual support, that pure dynamical models are genuinely explanatory, Kaplan and Bechtel (2011) counter-argued that appeal to these criteria fails to secure an explanatorily relevant status for pure dynamical models in cognitive science. Pure dynamical models, that is, ones that do not reference mechanisms, (re-)describe cognitive phenomena and may predict them, yet do not explain them. According to Kaplan and Bechtel (2011), bona fide explanation requires a specification of or reference to the underlying mechanisms that produce the behavioral regularities described in dynamical models. They thus claim, "Dynamical explanations do not provide a separate kind of explanation; when they explain phenomena, it is because they describe the dynamic behavior of mechanisms." (p. 440).

Indeed, the prevailing attitude seems to be that dynamical models in cognitive science (and neuroscience) explain only insofar or to the extent that they refer to details of mechanisms thought to produce, underlie, or otherwise be responsible for the phenomena described and targeted for explanation in dynamical models (Kaplan & Bechtel, 2011; Kaplan & Craver, 2011). I agree with the mainstream here: prediction, unification, and counterfactual support are by themselves not enough to secure an explanatory role for dynamical models.[5] Although dynamicists often stress the descriptive and predictive power of dynamical models (Stepp et al., 2011; Van Gelder, 1998; Van Gelder & Port, 1995), Kaplan and Bechtel (2011) rightly argue that predictive force is not enough for explanatory power. On the basis of law-like regularities, one may correctly predict the occurrence of events but fail to explain them. To use an often-rehearsed example, falling mercury in a barometer is a reliable indicator of impending storms, but mercury drops of course do not explain the occurrence of

storms. Rather, drops in atmospheric pressure are a common cause of – and so explain – both drops in the mercury values of barometers and the occurrence of storms. This of course was precisely the problem that was so devastating for Hempel's (1965) covering law model of explanation. So, by the same token, descriptive and predictive regularities are also not enough to secure *explanatory* counterfactual dependencies and unifying relations. Moreover, Kaplan and Bechtel (2011) point out that pure dynamical models often describe and relate independent and dependent variables or parameters at the macro-level of observable behaviors, and thus only describe effects (Cummins, 2000) and not the systemic features giving rise to these effects.[6] In short, such *predictivism* (Kaplan & Bechtel, 2011) falls short as an account of explanatory power.

In a recent paper, Gervais (2015) endorses this assessment, but argues that there is another source of explanatory power that has been overlooked in the debate, namely the ability to confer scientific understanding to scientists. Scientific understanding amounts to the ability of scientists to derive qualitative consequences (insights into how a system behaves or will behave in a certain way, that is, what it does or will do) from a given model without performing exact calculations.[7] Some pure dynamical models, by his lights, fare well on this score. His example concerns Voss' (2000) dynamical model of anticipating synchronization in circadian systems. Although an interesting take on the issue, this view also falls short in providing a suitable anchor for the explanatory power of pure dynamical models. Somewhat ironically, the reason is given by Gervais himself:

> As Kaplan and Bechtel stress (2011, p. 443), the Voss model is only a how-possible model: it has yet to be verified if the mammalian circadian system works that way … this hypothetical character has no bearing on the fact that the model conveys understanding … the realization that the mammalian system *might* operate in accordance with the Voss model. (2015, p. 63)

So as things stand now, the mammalian system might or might not operate in the way described in the Voss model. However, if plausibility, roughly, the likelihood that the model is correct, carries weight in explanatory power assessments, this approach fails: without reference to features that would convert a how-possible model into a how-actually model (be it information on the mapping between operations specified in dynamical models and their component parts, laws, causal relations, or something else that carries explanatory weight), dynamical models like the Voss model are merely hypothetical or how-possibly models (see Kaplan & Bechtel, 2011). Now, by their very nature, hypothetical models might be wrong. Hence, it is quite conceivable that they provide *incorrect understanding*: if one derives qualitative consequences from an incorrect model,

chances are good that these consequences are incorrect as well. But such derivations will most likely fail to provide genuine insight into the system under investigation.

This "plausibility worry" is accounted for by the perspective advanced in this paper in several ways: First, a wealth of behavioral evidence supports the claim that Thelen and colleagues' dynamical model of the A-not-B error (as well as the dynamical models of unintentional interpersonal coordination discussed in Section 5) tracks actual (causal) dependency relations. Second, although their dynamical model of the A-not-B error does not reference localization information with respect to the mapping of the operations specified in the dynamical model onto working parts, or information on the mechanisms constituting those operations, it is consistent with, *inter alia*, what is known about central nervous system function. In this sense, the model is evidentially supported by mechanistic evidence. Yet, such localization information is not what gives contextualized causal models their explanatory traction (I elaborate these points in Sections 3 and 4, where I also argue that there is a principled difference between such dynamical models and mechanistic models, be they "complete," schemata, or sketches. In a nutshell, such dynamical models list dependencies between external constraints and causal factors. Since such external constraints are not, on any reading of constitutive relevance, constituents of mechanisms, they are not articulated in mechanistic models. *Ipso facto*, dependencies between external constraints and core causal factors are not specified in mechanistic models. Therefore, contextualized causal models are importantly different from mechanistic models.)

In the next section, I elaborate a broadly interventionist perspective on the explanatory structure and power of some pure dynamical models in cognitive science, in casu Thelen and colleagues' (2001) dynamical model of the A-not-B error. I argue that this model can be understood as a "contextualized causal model." This explanatory structure gives such models genuine explanatory power.

3. Dynamical models as contextualized causal model explanations

3.1. *Contextualized causal model explanations*

Weber and colleagues (2013) argued that (some) dynamical models, among which Thelen and colleagues' (2001) model of the A-not-B error (and its variations) and the rhythmic finger movement model of Haken and colleagues (1985), are *causal* explanations, countering Walmsley's (2008) idea that (all) dynamical explanations are traditional covering law explanations. More specifically, Weber and colleagues (2013) argued that (some) dynamical models are non-deductive, *causal* covering law explanations, in which the "laws"

figuring in the explanans should be thought of as default rules, rather than exceptionless laws, because these "laws" do not cover each and every case (e.g., as regards perseverative reaching, it is usually the case that perseverative reaching occurs when infants have been repeatedly cued to a specific location; or, the model addresses most contextual subtleties that affect whether or not perseverative reaching occurs, yet not all contextual variation). The idea seems to be that the system of equations relating parameters and variables makes up a default rule or rules and that, by assigning specific values to these parameters and variables, one can derive specific reaching behaviors as a consequence of these interrelated parameter and variable values. Since these parameter and variable values describe (in mathematical fashion) causes of the explanandum (specific reaching behaviors can be derived as effects from these causes), the explanation is interpreted as causal. Weber and colleagues (2013) are on to an important idea, that is, that some dynamical models are causal ones, but much more can be said on the nature of the relations holding between the parameters and variables described in causal dynamical models. Specifically, such models, at least Thelen and colleagues' (2001) model of the A-not-B error and the models of unintentional interpersonal coordination developed by Richardson et al. (2007) and Schmidt and Richardson (2008), are *contextualized* causal models. That is, rather than taking every parameter and variable as corresponding to a cause simpliciter, such models capture dependencies between internal and external constraints and core causal factors, which specify when the core causal factor is (and when it isn't) a difference maker for the explanandum phenomenon. And, moreover, the specification of such dependencies affords formulating and answering explanatorily relevant what-if-things-had-been-different questions (see Sections 2.1 and 3.2).

Two issues require some clarification at this point – what view on causation and causal relations grounds the claim that these models are contextualized causal models and what makes these models relevantly different from mechanistic models. I address these issues in turn.

I endorse the view championed by Woodward (2003) and others (e.g., Craver, 2007; Hausman, 1998; Mackie, 1980) that causal relations are ones that are potentially exploitable for purposes of manipulation and control. If X is to count as a cause of Y it must be shown that X makes a difference to Y, namely, it should be possible to change Y by intervening on X, when other (off-path) variables in the system are held fixed to their values. The intervention on X thus should be such that the change in Y is only due to the change in X. Now, such "ideal" interventions are hard to come by in scientific practice or perhaps even impossible to attain. In scientific practice, it often seems not feasible to control for all of-path variables (known or unknown) that might affect a causal relation (as Woodward, 2003, 2017, p. 6, note 7 remarks, an intervention is an "idealized non-confounded experimental manipulation"; see also Craver, 2007). The best one can do, it seems, is to control for the

known and salient ones. The idea of causation and causal relations as difference making and difference making relations, respectively, clearly applies to the dynamical models discussed in this paper, and the developers of these models also make efforts to control for off-path variables. Let me explain.

Manipulations of the core causal factors cited in the models show that these make a difference to the explananda phenomena targeted by these models. For instance, manipulating the strength (value) of motor memory of previous reaches (variable) by administering repetitive cues over time to a certain location impacts the strength of the influence of motor memory on the decision to reach to a certain location, that is, perseverative reaching errors occur more often when infants receive more repetitive location cues. This is the "canonical" situation in which motor memory shows up as a key causal factor for perseverative reaching. But as Thelen and colleagues (2001) remark:

> The model offers a powerful and parsimonious, yet biologically plausible, account of the *many contextual influences* on A-not-B tasks that have puzzled developmental psychologists for two generations. (p. 2, emphasis added)

Thelen and colleagues' (2001) model is intricate, for it accounts both for the canonical situation and for contextual influences (internal and external constraints) that affect whether or not motor memory makes a difference to perseverative reaching. Changes to the values of these constraints show when the core causal factor is or isn't a difference maker for perseverative reaching. Like the relation between the core causal factor and the explanandum phenomenon, these contextual variations are also explicable in terms of an interventionist framework on causal relations.

In the canonical situation, the salient internal and external constraints are held fixed to a certain value and under those (fixed) conditions it is assessed whether or not motor memory makes a difference to perseverative reaching (by manipulating the number of location cues administered to the infants). We can think of these internal and external constraints held fixed to a certain value as off-path variables, that is, variables not on the causal path between motor memory and perseverative reaching. Furthermore, we can think of each of the contextual variations on or deviations from the canonical situation, captured by the model simulations, in which one of the internal or external constraints (off-path variables) is assigned a different value, as corresponding to a novel "causal scenario" in which it is again assessed whether motor memory (still) makes a difference to perseverative reaching. That is, in each novel causal scenario, an internal or external constraint is held fixed to a specific value different from the canonical situation (e.g., fixed to a specific different value with respect to the ambiguity of the task input, body posture of the infant, etc.), and it is assessed whether, under this condition, motor memory is still a difference maker for perseverative reaching or not. To think of these internal and external constraints

as off-path variables held fixed to a certain value in a particular causal scenario accords with the extensive simulation runs done with the model and also captures the insight that internal and external constraints are not to be thought of as causal factors on a causal (directed) path from constraint to core causal factor to explanandum phenomenon (effect), or directly from constraint to explanandum phenomenon (effect). For instance, constraints like body posture or ambiguity of the task input do not cause a specific value of motor memory of previous reaches, of course. Cued reaches to specific locations is what causes memory of these reaches. And neither do constraints directly cause perseverative reaching. For that to occur, memory of previous reaches has to be in place. These constraints rather set the context within which a core causal factor is a difference maker or not. In sum, the claim that Thelen and colleagues' model of the A-not-B error is a contextualized causal model can be backed up and made precise in terms of an interventionist framework on causal relations. Importantly, Thelen et al. (2001) and Smith and Thelen (2003) also clearly consider their model to be a causal explanation. The view that their model provides a causal explanation for the A-not-B error is a continuous thread running through their research. As Smith and Thelen (2003), for instance, state, the aim of the model "is to explain how the A-not-B error is the emergent product of multiple causes interacting over nested timescales … there are many causes that make the error appear and disappear" (pp. 345–346).[8] This is of a piece with their earlier mentioned claim that "infants make perseverative location errors because the motor memory of one reach persists and influences subsequent reaches" (Thelen et al., 2001, p. 9).

3.2. *Contextualized causal models are non-mechanistic*

Importantly, these dynamical models are causal, yet non-mechanistic. In a recent paper on abstraction and mechanistic explanation, Boone and Piccinini (2016) provide a generic characterization of mechanistic models that is widely endorsed in the literature: "At each level of organization, a mechanistic model articulates how the relevant properties [sets of causal powers] of the relevant components, suitably organized, produce the phenomenon" (2016, p. 4).

They also endorse the widely held view that mechanistic (model) explanations are *constitutive*. So, mechanistic models articulate how suitably organized *constitutively relevant* components and their properties constitute the phenomenon to be explained (one form of organization being the causal intra-level interactions between components. Inter-level relations between components and phenomena are constitutive and non-causal). Thus, whether a model articulates a mechanism in full constitutively relevant detail or abstracts away, for whatever reasons, from some or many constitutively relevant features, as in the case of mechanism sketches

and schemata, on this view, mechanistic models *only list constitutively relevant mechanistic features*.

Contextualized causal models, in contrast, are not mechanistic models (be they "complete," schemata, or sketches) since they, *inter alia*, list dependencies between external constraints and core causal factors. Such external constraints are not constituents of mechanisms (they fail mutual manipulability interpretations of constitutive relevance [Craver, 2007], they fail no-decoupling requirements on constitutive relevance [Baumgartner & Casini, 2017], etc.). Consider that two influential (and very different) mechanistic accounts of constitutive relevance – Craver's (2007) mutual manipulability account and Baumgartner and Casini's (2017) no-decoupling account – both require that evidence for constitutive relevance relations between parts (putative components) and phenomena entails that, following appropriate manipulations of a mechanism/mechanistic system, macro-level changes in phenomena must be accompanied by corresponding micro-level changes in components. If, following an appropriate manipulation, there is a macro-level change in a phenomenon without an accompanying change in the putative component, that putative component is not a constituent (genuine component) of the mechanism. Both these mechanistic accounts diverge greatly with respect to the nature of possible manipulations on mechanistic systems and the evidential import they convey, but both stress that evidence for the constitutive relevance of a specific part requires that when a macro-level phenomenon changes through an appropriate manipulation, the part must change as well.[9] In light of this requirement, we can easily see that external constraints are not, on such readings of constitutive relevance, constituents of mechanisms. For instance, if you manipulate body posture (for the sake of argument, a putative part), perseverative reaching (phenomenon) is changed as well. Yet, if you manipulate perseverative reaching (by engaging infants in the experimental task) body posture is not affected (e.g., in the canonical situation they do it sitting on the lap of an adult on different trials). Hence, body posture is not a constituent of perseverative reaching. Only should body posture change in each and every experimental condition would one have (abductive and fallible) evidence suggesting that body posture is a constituent in perseverative reaching. Likewise, if you manipulate the salience of the hiding conditions (again, for the sake of argument, a putative part), perseverative reaching is changed as well. Vice versa, however, when you manipulate perseverative reaching (by engaging infants in the experimental task), the salience of the hiding conditions is unaffected. The same goes, I submit, for each external constraint taken into account in the dynamical explanation by Thelen and colleagues (2001): these are not constituents of mechanisms.[10,11]

Moreover, external constraints are not to be confused with environmental or contextual conditions of a mechanism, at least not in the causal sense in which environmental/contextual conditions are understood in the mechanism literature. In the context of mechanistic explanation, it seems that environmental/

contextual features are stressed as explanatorily relevant either as causal background conditions that have a causal impact on mechanism function or as causally relevant components. For instance, how light understood as a causal background condition impacts – plays a causal role in initiating – the operation of the retina (Craver & Bechtel, 2007; Roe & Baumgaertner, 2017). However, as explained, external constraints are not to be thought of as causal factors on a causal (directed) path from constraint to core causal factor to explanandum phenomenon (effect), or directly from constraint to explanandum phenomenon (effect). Again, body posture does not cause a specific value of motor memory of previous reaches; cued reaches to specific locations do. Body posture, rather, is a constraint setting a context within which a core causal factor is a difference maker or not. External constraints, hence, are (also) different from environmental/contextual conditions of a mechanism as understood in the mechanism literature.

The distinction is important. What the above discussion tells us is that mechanistic (model) explanations track (inter-level) constitutive dependency relations and (intra-level) causal ones. Environmental/contextual conditions can figure in the latter dependency relations. External constraints, however, figure in neither: dependencies between external constraints and core causal factors are neither causal nor constitutive and, hence, not specified in mechanistic models (at least not according to extant conceptions of mechanistic models).[12] In light of this, we can also see that contextualized causal models have more explanatory power than mechanistic models with respect to explaining the intricacies of perseverative reaching (and, more generally, phenomena that are highly sensitive to contextual subtleties). Specifically, the model of Thelen and colleagues (2001) answers what-if-things-had-been-different questions by spelling out the mutual interdependencies between the core causal factor of motor memory of previous reaches and internal constraints, such as looking, motoric planning, and reaching, and external ones like the relative ambiguity of the task input and the delay between looking and reaching. Simulations are run with the model, informed by extensive empirical findings that characterize under which contextual conditions motor memory of previous reaches makes a difference to perseverative reaching and when it does not. For instance, as discussed earlier, how constraints like body posture, the delay between looking and reaching, and the salience of the hiding locations affect whether or not motor memory of previous reaches is a difference maker for perseverative reaching. These contextual dependencies are captured in terms of differential and difference equations, and these dependencies also tell us how changes to the values of the constraints (captured by mathematical parameters) in these dependencies result in changes in the value of the target explanandum phenomenon, that is, they tell us under which parametric value changes motor memory is and isn't a difference maker for perseverative reaching. In interventionist vein, Thelen and colleagues (2001) subscribe to this view by arguing that

action dynamics can be studied through behavioral manipulations: "we contend that the A-not-B error is … a window on these planning dynamics, with the ambiguous targets and delays providing the necessary *manipulations*" (Thelen et al., 2001, p. 13, emphasis added).

Mechanistic models, in contrast, do not articulate dependencies between external constraints and core causal factors and, hence, do not offer the resources to assess the contexts in which motor memory will or will not make a difference to the A-not-B error. That is, they do not offer the resources to answer what-if-things-had-been-different questions with respect to changes in these dependencies. This is precisely what contextualized causal models do[13] (although good mechanistic models do provide answers to other sorts of what-if-things-had-been-different questions, viz., questions with respect to possible and/or actual manipulations of mechanisms and their components [Craver, 2007]).[14]

Summing up, my view thus disagrees with Kaplan and Bechtel's (2011) perspective that dynamical explanations do not constitute a separate kind of explanation, distinct from mechanistic explanation. Rather, dynamical explanations, in fact, can provide a kind of explanation separate from mechanistic explanation and deliver important explanatory insights. In the case of the model of the A-not-B error, the model answers what if-things-had-been-different-questions, clarifying how the error and its decline depend on the nature of the contextual dependencies holding between motor memory of previous reaches and (a variety of) internal and external constraints.

There is a further important issue that emerges, of course: how do such dynamical models relate to mechanistic ones? I take the results of this section to establish that mechanistic information with respect to the mapping of the operations specified in contextualized causal models onto working parts, or their constitution by underlying mechanisms, is not needed to deliver the explanatory insights that contextualized causal models offer. That said, in the next section I argue that such mechanistic "mapping information" is relevant from an evidential point of view: it confers plausibility on the specification of operations in these contextualized causal models and, relatedly, information on mechanism manipulations is relevant to backing up the claim that the core causal factors listed in contextualized causal models indeed are difference makers for the targeted explananda phenomena.

4. Comparison: The evidential role of mechanistic information

Thelen and colleagues (2001) claim that their model is biologically plausible, despite the fact that it does not specify neural mechanisms. It is a model of the behavioral dynamics involved in perseverative reaching. Nevertheless, the dynamics specified by the model are in accord with

what is known about central nervous system function: the close couplings between looking, planning, motor memory, and reaching constitute processes that are also detected, through a large number of studies, in the brains of awake, behaving monkeys (Thelen et al., 2001, p. 9, table 2 marshals a great number of studies in support of this claim). These studies use tasks similar to the A-not-B error and suggest that motor cortical areas are recruited in sensorimotor transformations and in stimulus retention in memory. What's more, motor, premotor, parietal, and prefrontal cortical areas show increased activation in response to reaching tasks and variations in task conditions, similar to the A-not-B error task and its variations in task conditions.

Now, although one must be very cautious in extrapolating findings across species, these findings do suggest that the dynamics specified in the model explanation for perseverative reaching align with findings about the functioning of neural mechanisms. There thus is evidence suggesting that the dependencies between the internal constraints and the core causal factor captured in the model are produced by actual underlying mechanisms and that these mechanisms are responsive to (changes in) external constraints. This evidence thus confers plausibility, alongside the vast behavioral evidence on which the model is based, on the dynamics specified in the model explanation for perseverative reaching. But it is important to see that the fact that these processes (probably) correspond to component parts or underlying mechanisms is not what gives the model its explanatory traction. The key explanatory traction of contextualized causal models, rather, consists in showing how core causal factors make a difference to target explananda phenomena, relative to internal and external constraints. That is, these models answer what-if-things-had-been-different questions by capturing salient contextual variation under which the core causal factors are or aren't difference makers for explananda phenomena.

Such mechanistic information plays an evidential role, not an explanatory one. As I see it, the evidential import of mechanistic information, localization information in particular, comes in two (related) ways. First, it consists in showing that the operations specified in such contextualized causal models can be mapped onto working parts or are constituted by underlying mechanisms (see Piccinini & Craver, 2011). Such information thereby confers plausibility on contextualized causal models by indicating that the operations specified in them actually (or probably) are component processes of mechanisms or are constituted by mechanisms. Second, and relatedly, mechanistic information is evidentially relevant to backing up the claim that the core causal factors listed in contextualized causal models indeed are difference makers for the targeted explananda phenomena (they play this role alongside behavioral evidence). For instance, if the mechanism

constituting motor memory is incapacitated or weakened, this manipulation will impact perseverative reaching: it will then not occur, or not as strongly. So information on such mechanism manipulations confers plausibility on the claim that a core causal factor (indeed) is a difference maker. However, localization and manipulation information with respect to mechanisms is not required to *explain* the explananda phenomena, *and their contextual variability*, targeted by contextualized causal models. Rather, contextual dependencies are invoked to explain when a core causal factor is or isn't a difference maker for a target explanandum. So, contextualized causal models do the explanatory work, not localization and manipulation information with respect to mechanisms.

Another way to phrase this point is that models work at different levels or scales and address different questions; for example, what explains the error and its variations vs. what mechanism constitutes a core causal factor? For the latter explanation-seeking question, mechanistic models are explanatorily relevant (specific details of the explanatory request will further determine how much and what sort of localization information on working parts is required). Yet for the former explanation-seeking question, which of course is the question for which Thelen and colleagues (2001) developed their model, localization information concerning operations and their working parts or constituting mechanisms is not explanatorily relevant, since such information does not support answers to the relevant what-if-things-had-been-different questions. The model for the A-not-B error, however, precisely gives you these answers by articulating how motor memory makes a difference (or not) to perseverative reaching, relative to internal and external constraints. In sum, mechanistic information with respect to localization plays an evidential role, alongside behavioral evidence, in conferring plausibility on the behavioral processes specified in the dynamical model. But one can do without this information from an explanatory point of view.[15]

The upshot is that there are contexts in cognitive science in which pure dynamical explanations do provide a separate kind of explanation from mechanistic explanation, and in these contexts, when they explain phenomena, it is not because they describe dynamic mechanisms. In recent years, the point is increasingly being stressed that there are explanation-seeking contexts in the life sciences in which other types of explanation, like minimal model explanations (Batterman & Rice, 2014; Ross, 2015) and optimality explanations (Rice, 2015), are better suited than mechanistic explanations. The view developed here is on a par with these analyses.[16]

The "causal" analysis given in this paper is not restricted to dynamical models of the A-not-B error, but applies more broadly to other pure dynamical models in cognitive science, and thus strengthens the perspective that dynamical models can be explanatory and can have explanatory power despite

the fact that they do not describe mechanisms. In the next section I discuss a second example of dynamical explanation in cognitive science to substantiate this claim, namely dynamical explanations of *unintentional interpersonal coordination*, which are a recent extension of the HKB model of coordination dynamics (Richardson et al., 2007; Schmidt & Richardson, 2008). I close the paper in the final section by (briefly) contrasting my causal perspective on dynamical explanations with some recent analyses of dynamical explanation in systems neuroscience – which borders closely to cognitive science and according to some might be seen as a part of cognitive science – that also tell a positive story about the explanatory power of pure models, but differ from the perspective advanced here in a crucial respect. According to these analyses some dynamical models in systems neuroscience are *non-causal* minimal models. In contrast, the analysis advanced here is explicitly causal: causation and counterfactual dependence are a package deal. The more general lesson that I take to follow from these other works and the work discussed here is that the explanatory power of pure dynamical models – ones that do not refer to mechanisms at all – can be secured in different ways.

5. Interpersonal coordination dynamics

The influential HKB model of the bimanual coordination of rhythmic finger movements (Haken et al., 1985) concerns the *intrapersonal* coordination of movements. It subsequently inspired a wealth of research on *interpersonal* coordination (Richardson et al., 2007; Schmidt & Richardson, 2008), that is, coordinated movements between two people. As is well known, the HKB model specifies the rate of change of the relative phase angle of two oscillators (the left and right fingers) in terms of a differential equation that specifies the potential states in which the finger movements settle or remain. These states (and thus the relative phase angle of the oscillators) are essentially governed by parameters (specified in the equation) that reflect movement (oscillation) frequencies. Research indicates that rhythmic finger movements display two basic patterns: inphase (both fingers move simultaneously to the left or right in alternating fashion) or antiphase (one finger moves to the left, the other moves to the right). The patterns are dependent on oscillation frequencies; at low frequencies, both inphase and antiphase motion are observed. When the frequency of the movements increases, and a certain oscillation frequency is reached, only inphase motion is observed. The HKB model captures these behavioral patterns in mathematical fashion (the details of the equation need not concern us here).

5.1. *Unintentional interpersonal coordination*

Following up on these intrapersonal dynamics, the research by Richardson, Schmidt, and collaborators (2007, 2008) aims to investigate whether, and if

so, how, the coordination principles that govern intrapersonal movements apply to, and can be used to study, the coordination principles that govern interpersonal coordinated movements. A wealth of research indicates that this indeed is the case, both for intentional coordinated movements between two people and unintentional coordinated movements. These movements all display the same behavioral patterns captured and predicted by the HKB model.

Unlike intentional interpersonal coordination, in *unintentional* interpersonal coordination subjects are *not* instructed to coordinate their movements. They are instructed to produce rhythmic movements at their own preferred pace in proximity to one another, such as wrist pendulum-movements, but they are asked to do so in light of other research objectives than studying movement coordination, like studying aspects of problem solving or assessing ergonomic features of artifacts, such as rocking chairs (Richardson et al., 2007). However, despite the fact that subjects are not instructed to coordinate their movements, and despite the fact that they are – as debriefings afterwards indicate – not aware of the true purpose of the experiments, which of course is studying coordination phenomena, interpersonal coordination of movements does take place in such experiments.

A case in point is the research by Richardson and colleagues (2007) on unintentional interpersonal coordination of rocking chair movements. They investigated whether coordinated rocking chair movements also emerge when subjects are not instructed to coordinate their movements and, if so, whether the manipulation of visual focus (sharp visual focus of the other subjects' movements, observing the other subjects' movements from the periphery, or no visual information available) makes a difference to unintentional interpersonal coordination. In each trial, two subjects were instructed to rock a chair at their own preferred pace and were told that the experiment was devised to investigate certain ergonomic features of the rocking chairs. Subjects were furthermore (casually) told that they were performing the rocking task in pairs for the sake of efficiency considerations in data collection. They were also informed that the experiment was conducted to assess how different postural configurations have an impact on the stability of the rocking chairs. This was "tested" in different trials by having the subjects look, while rocking, at a red dot at the arm rest of the chair of the other subject (sharp visual focus), at a red dot right in front of them (peripheral visual focus), or at a red dot to the side of them (no visual information).

The results indicate that in this experimental setting, visual focus is a core causal factor that makes a difference to unintentional interpersonal coordination, and does so in different ways relative to specific external constraints. One salient external constraint concerns mass differences between the rocking chairs (which were manipulated by the experimenters

by adding a 27 kg mass to the base of some chairs), leading to different natural "rocking periods" of the chairs.[17] As expected, in the no visual information condition (in which subjects did not see each other's movements) entrainment or coordination of rocking movements did not occur. However, in both the central and peripheral vision conditions, in line with the HKB model, unintentional interpersonal coordination did occur: relative phase angles of the oscillators (rocking chair movements) near the attractor states (inphase and antiphase) became dominant when subjects could see one another's movements, despite the fact that they were not instructed to coordinate their movements. In other words, visual focus makes a difference to the phenomenon of unintentional interpersonal coordination. Interestingly, it does so in different ways relative to the external constraint of mass difference, and thus different natural rocking periods, between the rocking chairs.

Specifically, central visual focus makes a difference to the unintentional interpersonal coordination of movements (measured in terms of distributions of the relative phase angles of the oscillators during the 90s trials) both when there are 27 kg mass differences between the rocking chairs and when there are no mass differences. However, central visual focus has a greater impact on the unintentional interpersonal coordination of movements when the chairs have the same mass than when they have a different mass. Interestingly, in the peripheral vision condition, visual focus does not make a difference across different mass conditions. It rather only makes a difference to the unintentional interpersonal coordination of movements when the chairs have the same mass. When the chairs differed in mass by 27 kg, unintentional interpersonal coordination of movements did not take place.

5.2. *Contextualized causal model explanations, again*

So what we see here is that this research can be interpreted as advancing a contextualized causal model explanation of the unintentional interpersonal coordination of rocking movements in which visual focus is the core causal factor that makes a difference to the phenomenon. Furthermore, constraints are listed that affect the manner in which visual focus makes a difference: the impact it has on the explanandum phenomenon is stronger in the central vision condition than in the peripheral vision condition. Furthermore, vision in the central condition makes a difference to coordinated movements across different mass conditions in different ways: the effect is stronger when the chairs have equal mass than when they differ in mass. Finally, in the peripheral vision condition, visual focus only makes a difference in the no mass difference condition. When chairs differ in mass, (peripheral) visual focus is not a difference maker for unintentional

interpersonal coordination. In other words, contextual dependencies between the core causal factor and constraints are given that specify when the core causal factor is (and when it isn't) a difference maker for the explanandum phenomenon, and how strong an effect it has on the phenomenon. These dependencies thus answer what-if-things-had-been-different questions, for they tell us under which conditions, and changes therein, the core causal factor makes a difference to the explanandum phenomenon, how strong its effect is, and when it does not make a difference. As argued at length, the explanatory power of contextualized causal model explanations resides in this ability.

Furthermore, like the model of the A-not-B error, these models are "pure" in the sense that they do not describe underlying mechanisms, and need not do so in order to confer relevant explanatory insights. The developers of the model make this explicit by stating that:

> Rhythmic interlimb coordination appears to be the result of the lawful relations that exist between the sub-components [oscillators, such as moving fingers, legs, and wrists] of perceptual-motor systems, rather than a specific anatomical or neural mechanism. (Richardson et al., 2007, p. 869)

To be sure, the operation of neural mechanisms is, of course, required for agents to engage in oscillatory behaviors. The point is that such mechanisms need not be referred to in explanatory models of (intrapersonal and) interpersonal coordination. Like the model of the A-not-B error, the model of unintentional interpersonal entrainment captures different contexts under which the core causal factor is or isn't a difference maker for the target explanandum. For this explanatory request, the articulation of dependencies between contextual constraints and core causal factors is what matters. In the case at hand, answers require specifying different values of the behavioral and contextual parameters, such as oscillation frequencies, and visual focus and mass difference conditions. This is precisely what can be done with contextualized causal dynamical models of unintentional interpersonal coordination. As argued at length in Section 4, mechanistic information with respect to, *inter alia*, localization is evidentially relevant, but it does not give explanatory traction in the explanatory context for which the model of unintentional interpersonal entrainment was developed, namely identifying the key causes of unintentional interpersonal entrainment and the contextual variation it is subject to.

6. Discussion and conclusions

In this paper I have argued that some "pure" dynamical models – ones that do not refer to mechanisms at all – in cognitive science are "contextualized causal models" and that this explanatory structure gives such models genuine explanatory power. I built this analysis in terms of dynamical

models of the A-not-B error in developmental psychology (Thelen et al., 2001) and dynamical models of unintentional interpersonal coordination (Richardson et al., 2007). I contrasted this view with several other perspectives on the explanatory power of pure dynamical models. One of the main insights is that dynamical models need not refer to underlying mechanisms in order to be explanatory.

The account elaborated in this paper agrees with some other recent analyses of dynamical explanations in the context of systems neuroscience as regards the idea that dynamical models can have explanatory value despite the fact that they do not describe mechanisms (Chirimuuta, 2017; Ross, 2015; Woodward, 2017), but it is important to carefully distinguish these perspectives. Whereas the account advanced here is explicitly causal – causation and counterfactual dependence go hand in hand – these other perspectives argue that some dynamical models in systems neuroscience are *non-causal* minimal models (some of which explicitly prize causation and counterfactual dependence apart [Chirimuuta, 2017]). In my view, both perspectives are valuable and complementary. Consider that the explanatory traction of contextualized causal dynamical models hinges on capturing *variation*, that is, spelling out under which conditions a core causal factor is or isn't a difference maker for the explanandum phenomenon. Minimal models, and thus dynamical models plausibly regimented as minimal models, in contrast, serve an opposite explanatory goal. These models serve to explain how physically distinct systems all display the same, *uniform* macro-level behaviors, and do so by identifying those (non-causal) features that are common to these systems and in virtue of which they are able to display uniform macro-level behaviors. For instance, Ross (2015) discusses neuroscientists' explanations of why distinct neural systems exhibit the same qualitative patterns of neural excitability – how depolarization and repolarization trajectories change over time – in terms of the idea that the application of mathematical abstraction techniques, invoked to abstract away from details of mathematical models of specific neural systems, all produce equivalent abstracted models, that is, they converge to a single "canonical" model. Such abstraction procedures indicate that system-specific mechanistic features are explanatorily irrelevant for explaining uniform macro-level behaviors of these systems. What, rather, is explanatorily relevant are the common features across the abstracted dynamical models, that is, the features listed in the canonical model, such as the topological changes that neurons display if they transition from resting states to sustained firing states.

The more general lesson that I take to follow from these works on dynamical minimal models and the "causal" work discussed here is that the explanatory power of "pure" dynamical models – ones that do not refer to mechanisms at all – can be secured in different ways, and that the

structure and value of these models cannot be seen in isolation from the explanatory ends for which they are used.

Notes

1. Stepp and colleagues (2011) advance multiple criteria for the explanatory power of dynamical models, namely predictive power, counterfactual support, and unification. In Section 2.2, I argue that this is not enough to secure the explanatory power of such models.
2. Let me note that I restrict my account to the explanatory value of dynamical models in cognitive science. Furthermore, I purposely claim that *most* extant defenses of the explanatory power of pure dynamical models in cognitive science fall short of securing the explanatory value of such models. If one is willing to count systems neuroscience as part of cognitive science then, by my lights, there are some promising accounts available as regards the explanatory power of pure dynamical models. In the context of systems neuroscience, I take several authors to have convincingly argued that non-mechanistic dynamical models also can have explanatory force in virtue of these models being minimal models in the sense discussed by Batterman (2002) and Batterman and Rice (2014) (Chirimuuta, 2017; Ross, 2015; Woodward, 2017). That said, these accounts differ in a crucial respect from the one elaborated here: they treat dynamical models as non-causal ones. In contrast, the account advanced in this paper is explicitly causal. I briefly contrast my account with minimal model interpretations of dynamical models in Section 6. One of the important insights that derive from this comparison is that the explanatory power of non-mechanistic dynamical models can be secured along both causal and non-causal lines.
3. One of the reasons for assessing the model of the A-not-B error in detail is that its explanatory structure (and thus its explanatory power) has been misrepresented in the literature. For instance, while some take the model to be an instance of covering law explanation (Walmsley, 2008), others interpret the model in mechanistic fashion (Zednik, 2011). Yet others have been silent about its structure and power while invoking the model as a paradigmatic case of dynamical explanation (Van Gelder, 2006). Another reason is historical: it is considered one of the flagship cases of successful dynamical explanation in cognitive science. I also discuss more recent work on dynamical explanation in cognitive science in order to show that the "contextualized causal model" interpretation advanced in this paper is not restricted to the model of the A-not-B error, but applies more broadly to dynamical explanations in cognitive science.
4. As said, I take constraints to refer to task conditions and differentiate between internal and external constraints. External constraints refer to specific characteristics or properties of the task environment (e.g., the ambiguity of the task input in the A-not-B task) under which a task is (to-be) performed by subjects (e.g., infants in the A-not-B task). Internal constraints refer to specific restrictions on the behaviors (e.g., the specific delay between looking and reaching in the A-not-B task) that subjects (e.g., infants in the A-not-B task) (are to) execute during the task. These constraints are represented as parameters (and not as variables) in the dynamical models considered in this paper.
5. Of course, in Woodward's account explanation also involves predictive elements: answering what-if-things-had-been-different questions also concerns predicting

what would happen in counterfactual scenarios. As will become clear in this section, the claim that prediction does not provide sufficient grounds for explanatory power concerns the idea that law-like regularities, while having predictive credentials, by themselves need not be explanatory. The claim is not made in reference to Woodward's account, in which a variety of additional constraints are imposed on generalizations being truly explanatory.

6. Parameters refer to conditions that affect or constrain how dynamical behavior unfolds in real time, for example, how motor memory of previous reaches impacts perseverative reaching. Dynamical behavior, such as perseverative reaching behavior, is captured by means of variables that are assigned values which change from one time step to the next, for example, the hand reaching trajectory of an infant during the A-not-B task.
7. This skill-based account of understanding originates from the contextual theory of understanding, which was originally developed by De Regt and Dieks (2005) in the context of scientific theories in physics. Gervais (2015) applies the account to Voss' dynamical model.
8. Smith and Thelen (2003) speak about "multiple causes"; I prefer to speak about a core causal factor (i.e., motor memory) and internal and external constraints that affect whether or not the core causal factor is a difference maker for perseverative reaching. This re-characterization more precisely captures what goes on in the experimental investigations and is in line with the repeated emphasis on motor memory as cause of the A-not-B error.
9. Whereas Craver (2007) stresses ideal interventions that induce changes in one level by inducing changes in the other, Baumgartner and Casini (2017) argue that such ideal interventions, as defined by Craver (2007), are impossible in principle when the relation between micro and macro levels is one of constitution. In their view, micro and macro levels, in the case of constitution, can only be manipulated via *fat-handed* interventions that cause changes at both levels via separate causal paths. These differences, although intriguing, need not concern us here.
10. A similar argument can be run as regards internal constraints: since these refer to restrictions on subjects' behaviors that are (to be) executed during tasks (e.g., the specific delay between looking and reaching in the A-not-B task) and not to such behaviors themselves, such task conditions are also not constituents of mechanisms. I focus the discussion on external constraints since I take their non-constitutive (and non-causal) nature to be especially salient and establishing that suffices for my purposes here, namely demonstrating that mechanistic and contextualized causal models are importantly different.
11. Contextualized causal models are (also) different from mechanistic models as construed by Woodward (2013). Woodward considers mechanisms to be sets of (modular) causal relationships. That is, he takes mechanisms to consist of components/parts that can be characterized by variables, where these variables stand in causal (i.e., difference making) relationships to one another, and in which these relationships are viewed as intermediate or intervening links along the causal paths connecting mechanisms' overall input to their overall output (these overall input-output relationships characterize the overall behavior of mechanisms, i.e., their phenomena). Mechanistic models represent such sets of (modular) causal relationships. Now, Woodward says much more about features of mechanisms and models thereof (in particular, with respect to stability, modularity, and sensitivity to organization), but this brief sketch suffices to see that mechanistic models as understood by Woodward are not to be equated with contextualized causal models: the latter do

not represent sets of (modular) causal relationships. As said in Section 3.1, internal and external constraints listed in contextualized causal models are not to be thought of as causal factors on a causal (directed) path from constraint to core causal factor to explanandum phenomenon (effect), or directly from constraint to explanandum phenomenon (effect). For instance, constraints like body posture or ambiguity of the task input do not cause a specific value of motor memory of previous reaches, of course. Cued reaches to specific locations is what causes memory of these reaches. And neither do constraints directly cause perseverative reaching. For that to occur, memory of previous reaches has to be in place. These constraints rather set the context within which a core causal factor is a difference maker or not. So, contextualized causal models articulate contextual dependencies between constraints and core causal factors, not causal relationships between them (the causal relationship holds between core causal factors and target explananda, relative to these constraints).

12. I am not inclined to stretch the concept of mechanistic model further to also include external constraints. There then would be little distinctive left about such models; they may then include virtually anything. One needs to draw the line somewhere.
13. Schöner and co-workers have, since Thelen and colleagues' (2001) publication, focused a lot on dynamic field models understood as dynamic *neural* field models in which "the different factors that impact on behavior are conceived of as 'forces' in neural dynamics, ranging from intrinsic factors that reflect the neural circuitry to environmental factors that act through sensory input. The joint effect of these forces is the emergence of a stable state that becomes visible as overt behavior" (Maruyama, Dineva, Spencer, & Schöner, 2014). The dynamic neural field model, in the context of the A-not-B task, represents reaching directions in terms of activation levels of neurons, that is, in terms of the distribution of neural activation in the field. Such neural activation distributions results from perceptual inputs (specifics of the task input) and recent motor memory traces. This might suggest that the dynamic neural field model captures all the features of the original dynamic field model of the A-not-B error in terms of characterizations of neural activations. This is not the case, however. The model only articulates reaching directions and motor memory traces in terms of characterizations of neural activations. Specifics of the task environment (i.e., the external constraints) are considered key to infant perseverative reaching and are viewed as "forces" providing input to these neural activations, but are not themselves considered to be neural activation patterns and neither are they represented as such (Maruyama et al., 2014). In light of this, it would be a mistake to consider dynamic *neural* field models of the A-not-B error to be merely mechanistic models of neural mechanisms underlying the A-not-B error, for external constraints are assigned key importance. What it does show is that since the 2001 formulation, work has been done on the neural underpinnings of the internal operations listed in the 2001 model, thus conferring plausibility on the dynamics specified in the 2001 model (see also Section 4).
14. A worthwhile follow-up project would be to assess what an application of Hitchcock and Woodward's (2003) (comparative) account of explanatory power, which builds upon Woodward's (2003) account of causal explanation, to contextualized causal models would look like. For Woodward (e.g., 1997, 2000, 2003), the capacity to answer what-if-things-had-been-different questions is a necessary condition for an explanation to have genuine explanatory import or explanatory power. I take the explanatory power of contextualized causal models to also reside in this ability (see Sections 2.1 and 3.2). Hitchcock and Woodward (2003) also argue that one

explanation can do better than another one with respect to this capacity – that one explanation may be able to answer more what-if-things-had-been-different questions than another one and in this sense has more explanatory power than its counterpart. I also think that this is a fruitful way to assess the comparative explanatory power of contextualized causal models vis-à-vis mechanistic models. Hitchcock and Woodward (2003) cash out this idea of comparative explanatory power in terms of one explanatory generalization being more invariant under (testing) interventions than another one; the more invariant explanatory generalization will answer more what-if-things-had-been-different questions than its less invariant counterpart. In general terms, a generalization that describes a causal dependency relationship between explanans and explanandum variables is invariant if it would continue to hold – remain stable or unchanged – if various other conditions were to change. However, whereas invariance under interventions and the capacity to answer what-if-things-had-been-different questions are a package deal in Woodward and Hitchcock's (comparative) account of explanatory power, we need to prize these features apart in the context of contextualized causal models. To see this, consider that the dependency relation articulated in Thelen and colleagues' (2001) model between motor memory and perseverative reaching is a fragile one; whether or not motor memory is a difference maker for perseverative reaching is relative to a variety of contextual constraints – manipulate these constraints (or off path variables) and the dependency relation gets affected as well. Understanding these contextual subtleties of the A-not-B error is precisely what drove Thelen and colleagues' (2001) research into the phenomenon. The power of the model precisely resides in this feature of making explicit a number of constraints under which motor memory is and isn't a difference maker for perseverative reaching. So it answers relevant what-if-things-had-been-different questions by highlighting the context-sensitive – fragile – nature of the dependency relationship between motor memory and the A-not-B error. The capacity to answer what-if questions and invariance under interventions here, in a sense, pull in opposite directions.

15. Similar views can be found in Glennan (2005) and Woodward (2017) with respect to how much detail ought to be included in mechanistic models. Woodward, for instance, in discussing the Hodgkin and Huxley (HH) model of the action potential, writes: "The HH model shows that the generation of the action potential depends on (or requires at a minimum), among other things, the existence of at least two voltage gated and time dependent ion channels … given that such a structure is present and behaves appropriately, the presence of the specific mechanism by which the ion channels in the giant squid operates is not required for [explaining the] the generation of the action potential, as long as some mechanism or other that plays this role is present." (2017, p. 28).
16. To be sure, I do think that there are a lot of explanatory contexts in the life sciences where mechanistic explanations do provide the best explanations.
17. Although the weight of a person sitting in a rocking chair decreases the chair's natural period by elevating the center of mass of the chair, different weights of the subjects in other rocking chair experiments did not have a significant effect on movement coordination. Although weight differences were not recorded in the current experiment, given these previous outcomes, the experimenters take it that possible effects of weight differences in all likelihood are minimal.

Acknowledgements

I thank Raoul Gervais for useful comments. I am particularly indebted to two referees who provided very constructive feedback on previous versions of this paper. I also wish to thank Bill Bechtel for relevant suggestions.

Disclosure statement

No potential conflict of interest was reported by the author.

Notes on contributor

Dingmar van Eck is postdoctoral researcher in philosophy of science at the Centre for Logic and Philosophy of Science, Ghent University. Most of his current research is on issues related to scientific explanation in the life sciences and engineering sciences.

References

Batterman, R. (2002). Asymptotics and the role of minimal models. *The British Journal for the Philosophy of Science*, *53*(1), 21–38.

Batterman, R., & Rice, C. (2014). Minimal model explanations. *Philosophy of Science*, *81*(3), 349–376.

Baumgartner, M., & Casini, L. (2017). An abductive theory of constitution. *Philosophy of Science*, *84*(2), 214–233.

Bechtel, W. (1998). Representations and cognitive explanations: Assessing the dynamicist's challenge in cognitive science. *Cognitive Science*, *22*(3), 295–318.

Boone, W., & Piccinini, G. (2016). Mechanistic abstraction. *Philosophy of Science*, *83*, 1–13.

Chirimuuta, M. (2017). Explanation in computational neuroscience: Causal and non-causal. *The British Journal for the Philosophy of Science*. Advance online publication.

Craver, C. (2007). *Explaining the brain*. Oxford: Clarendon Press.

Craver, C. F., & Bechtel, W. (2007). Top-down causation without top-down causes. *Biology and Philosophy*, *22*, 547–563.

Cummins, R. (2000). "How does it work?" vs. "What are the laws?". Two conceptions of psychological explanation. In F. Keil & R. Wilson (Eds.), *Explanation and cognition* (pp. 117–145). Cambridge, MA: MIT Press.

De Regt, H., & Dieks, D. (2005). A contextual approach to scientific understanding. *Synthese*, *144*, 137–170.

Gervais, R. (2015). Mechanistic and non-mechanistic varieties of dynamical models in cognitive science: Explanatory power, understanding, and the 'mere description' worry. *Synthese*, *192*, 43–66.

Gervais, R., & Weber, E. (2011). The covering law model applied to dynamical cognitive science: A comment on Joel Walmsley. *Minds and Machines*, *21*, 33–39.

Glennan, S. (2005). Modeling mechanisms. *Studies in the History and Philosophy of the Biological and Biomedical Sciences*, *36*(2), 375–388.

Haken, H. L., Kelso, J. A. S., & Bunz, H. (1985). A theoretical model of phase transitions in human hand movement. *Biological Cybernetics*, *51*, 347–356.

Hausman, D. M. (1998). *Causal asymmetries*. Cambridge: Cambridge University Press.

Hempel, C. (1965). Aspects of scientific explanation. In C. Hempel (Eds.), *Aspects of scientific explanation and other essays in the philosophy of science* (pp. 331–496). New York: Free Press.

Hitchcock, C., & Woodward, J. (2003). Explanatory generalizations, part 2: Plumbing explanatory depth. *Noûs*, *37*, 181–199.

Kaplan, D., & Craver, C. (2011). The explanatory force of dynamical and mathematical models in neuroscience: A mechanistic perspective. *Philosophy of Science*, *78*, 601–627.

Kaplan, D. M., & Bechtel, W. (2011). Dynamical models: An alternative or complement to mechanistic explanations? *Topics in Cognitive Science*, *3*, 438–444.

Mackie, U. (1980). *The cement of the universe*. Oxford: Clarendon Press.

Maruyama, S., Dineva, E., Spencer, J. P., & Schöner, G. (2014). Change occurs when body meets environment: A review of the embodied nature of development. *Japanese Psychological Research*, *56*(4), 385–401.

Piaget, J. (1954). *The construction of reality in the child*. New York: Basic Books.

Piccinini, G., & Craver, C. F. (2011). Integrating psychology and neuroscience: Functional analyses as mechanism sketches. *Synthese*, *183*, 283–311.

Rice, C. (2015). Moving beyond causes: Optimality models and scientific explanation. *Noûs*, *49*, 589–615.

Richardson, M. J., Marsh, K. L., Isenhower, R. W., Goodman, J. R. L., & Schmidt, R. C. (2007). Rocking together: Dynamics of intentional and unintentional interpersonal coordination. *Human Movement Science*, *26*, 867–891.

Roe, R. M., & Baumgaertner, B. (2017). Extended mechanistic explanations: Extending the current conception to include more complex biological systems. *Journal for General Philosophy of Science*. Advance online publication.

Ross, L. N. (2015). Dynamical models and explanation in neuroscience. *Philosophy of Science*, *82*, 32–54.

Schmidt, R. C., & Richardson, M. J. (2008). Dynamics of interpersonal coordination. In A. Fuchs & V. K. Jirsa (Eds.), *Coordination: Neural, behavioral, and social dynamics* (pp. 281–308). Berlin: Springer-Verlag.

Silberstein, M., & Chemero, A. (2013). Constraints on localization and decomposition as explanatory strategies in the biological sciences. *Philosophy of Science*, *80*, 958-970.

Smith, L. B., & Thelen, E. (2003). Development as a dynamic system. *Trends in Cognitive Sciences*, *7*(8), 343–348.

Stepp, N., Chemero, A., & Turvey, M. T. (2011). Philosophy for the rest of cognitive science. *Topics in Cognitive Science*, *3*, 425–437.

Thelen, E., Schöner, G., Scheier, C., & Smith, S. B. (2001). The dynamics of embodiment: A field theory of infant perseverative reaching. *Behavioral and Brain Sciences*, *24*, 1–34.

Van Gelder, T., & Port, R. F. (1995). It's about time: An overview of the dynamical approach to cognition. In R. F. Port & T. Van Gelder (Eds.), *Mind as motion* (pp. 1–43). Cambridge, MA: MIT Press.

Van Gelder, T. (1998). The dynamical hypothesis in cognitive science. *Behavioral and Brain Sciences*, *21*, 615–628.

Van Gelder, T. (2006). Revisiting the dynamical hypothesis in cognitive science. In P. Farias & J. Queiroz (Eds.), *Advanced issues on cognitive science and semiotics* (pp. 73–91). Aachen: Shaker Verlag, Aachen.

Voss, H. U. (2000). Anticipating chaotic synchronization. *Physical Review E*, *61*(5), 5115–5119.

Walmsley, J. (2008). Explanation in dynamical cognitive science. *Minds and Machines*, *18*, 331–348.

Weber, E., Van Bouwel, J., & De Vreese, L. (2013). *Scientific explanation*. Dordrecht: Springer.

Woodward, J. (1997). Explanation, invariance, and intervention. *Philosophy of Science*, *64*, S26–S41.
Woodward, J. (2000). Explanation and invariance in the special sciences. *British Journal for the Philosophy of Science*, *51*, 197–254.
Woodward, J. (2003). *Making things happen. A theory of causal explanation*. New York: Oxford University Press.
Woodward, J. (2013). Mechanistic explanation: Its scope and limits. *Proceedings of the Aristotelian Society Supplementary Volume*, *87*, 39–65.
Woodward, J. (2017). Explanation in neurobiology: An interventionist perspective. In D. Kaplan (Ed.), *Explanation and integration in mind and brain science*. Oxford University Press.
Zednik, C. (2011). The nature of dynamical explanation. *Philosophy of Science*, *78*, 238–263.