COMPUTATIONALISM, DYNAMICAL SYSTEMS MODELS, AND INTERACTIVISM: A METHODOLOGICAL COMPARISON USING LAKATOS'S SCIENTIFIC RESEARCH PROGRAMMES APPROACH

HAYDAR OĞUZ ERDİN

BOĞAZİÇİ UNIVERSITY

2021

COMPUTATIONALISM, DYNAMICAL SYSTEMS MODELS, AND INTERACTIVISM: A METHODOLOGICAL COMPARISON USING LAKATOS'S SCIENTIFIC RESEARCH PROGRAMMES APPROACH

Thesis submitted to the

Institute for Graduate Studies in Social Sciences

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Philosophy

by

Haydar Oğuz Erdin

Boğaziçi University

2021

DECLARATION OF ORIGINALITY

I, Haydar Oğuz Erdin, certify that

- I am the sole author of this thesis and that I have fully acknowledged and documented in my thesis all sources of ideas and words, including digital resources, which have been produced or published by another person or institution;
- this thesis contains no material that has been submitted or accepted for a degree or diploma in any other educational institution;
- this is a true copy of the thesis approved by my advisor and thesis committee at Boğaziçi University, including final revisions required by them.

Signature....................OL....................

Date ....12/01/2021....................

ABSTRACT

Computationalism, Dynamical Systems Models, and Interactivism:

A Methodological Comparison

Using Lakatos's Scientific Research Programmes Approach

This thesis gives a methodological appraisal of the major research paradigms in cognitive science by making their heuristics explicit. The way their heuristics is made explicit and the appraisals given is based on Lakatos's methodology of scientific research programmes. The emphasis, however, is not on empirical progress but on heuristic type. Assuming that all the methodologies considered have equal empirical progress, it is argued that their heuristics differ when it comes to Lakatos's "continuity" and "autonomy" norms. This shows that computationalism and interactivism are methodologically on a par and have the strongest heuristics, whereas pure dynamical modelling has the weakest. Radical embodied cognitive science's model-based approach has medium heuristic strength, whereas the theory-based approach has weak heuristic strength. The thesis demonstrates that Lakatos's philosophy of science is relevant and effective when it comes to methodological issues in cognitive science. Since its finer distinctions reveal methodological constrains that are usually missed in the current debates regarding explanation and methodology in philosophy of cognitive science, it must be considered as an essential addition to the existing approaches to these issues. An attempt to modelling heuristic types based on the interactivist model of autonomy is also provided.

ÖZET

Bilgisayımsalcılık, Dinamik Sistemler Modelleri ve Etkileşimselcilik:

Lakatos’un Bilimsel Araştırma Programları Yaklaşımını Kullanarak

Yöntembilimsel Bir Karşılaştırma

Bu tez bilişsel bilimdeki önemli araştırma paradigmalarının sezgisel yöntemlerini belirginleştirerek yöntembilimsel bir değerlendirmelerini yapar. Bu paradigmaların sezgisel yöntemlerinin belirginleştirilmesi ve değerlendirilmesi Lakatos’un bilimsel araştırma programlarının yöntembilimine dayanmaktadır. Ağırlık görgül ilerlemeye değil, sezgisel yöntemin türüne verilmiştir. İncelenen yöntembilimlerin görgül ilerlemesinin aynı olduğu varsayılarak sezgisel yöntemlerinin Lakatos’un “süreklilik” ve “özerklik” normları açısından farklılıklar gösterdikleri savunulmuştur. Bu analiz bilgisayımsalcılık ve etkileşimselciliğin yöntembilimsel olarak aynı düzeyde olduklarını ve en güçlü sezgisel yönteme sahip olduklarını göstermiştir. Buna karşılık salt dinamik modellemenin en kötü sezgisel yönteme sahip olduğu ortaya konmuştur. Radikal bedenlenmiş bilişsel bilimin modellemeye dayalı yaklaşımının orta güçte sezgisel yönteme sahip olduğu ama teoriye dayalı yaklaşımının zayıf bir sezgisel yönteme sahip olduğu savunulmuştur. Bu tez Lakatos’un bilim felsefesinin bilişsel bilimdeki yöntembilimsel meseleler için yerinde ve etkili olduğunu göstermiştir. Bu yaklaşımın sağladığı daha yüksek çözünürlükteki ayırımlar, bilişsel bilim felsefesindeki açıklama ve yöntembilime dair güncel tartışmalarda genellikle gözden kaçırılan yöntembilimsel sınırlamaları ortaya çıkardığı için, halihazırda var olan yaklaşımlara önemli bir katkı olarak dikkate alınmalıdır. Etkileşimselciliğin özerklik modelini kullanarak, sezgisel yöntem türlerinin modellenmesine de çalışılmıştır.

CURRICULUM VITAE

NAME: Haydar Oğuz Erdin

DEGREES AWARDED

PhD in Philosophy, 2021, Boğaziçi University

Pedagogical Education Programme, 2018, Marmara University

MA in Philosophy, 2012, Bilgi University

MA in Mathematics, 2006, University of Texas at Austin

BA in Mathematics, 2000, Middle East Technical University

AREAS OF SPECIAL INTEREST

Philosophy of cognitive science, interactivism, ecological psychology, methodology of scientific research programmes, dynamical systems, emergence, process metaphysics

PROFESSIONAL EXPERIENCE

Teaching Assistant, FDP, Sabancı University, 2020 – present

Research Scholar, Tübitak Project 114K348, 2014-2017

Instructor, Mathematics Department, UT at Austin, 2008-2009

Teaching Assistant, Mathematics Department, UT at Austin, 2002-2008

AWARDS AND HONORS

2nd Place, Graduate School of Social Sciences, Bilgi University, 2012

GRANTS AND SCHOLARSHIPS

Boğaziçi University BAP-D Project 18B02D1, 2018-2020

Tübitak 1001 Project 114K348, 2014-2017

Bilgi University Graduate Scholarship, 2010-2012

Research Assistant Grant, UT at Austin, 2006

David Bruton, Jr. Graduate Fellowship in Mathematics, 2002-2003

PUBLICATIONS

Journal Articles

Erdin, H.O. (2020). Appraisal of Certain Methodologies in Cognitive Science Based on Lakatos's Methodology of Scientific Research Programmes. *Synthese S.I.:*

Explanations in Cognitive Science: Unification vs Pluralism. https://doi.org/10.1007/s11229-020-02612-4

Erdin, H.O., and Bickhard, M. (2018). Representing is something that we do, not a structure that we “use”: Reply to Gładziejewski. *New Ideas in Psychology*, *49*, 27-37.

Master Theses

Erdin, H. O. (2012). Žižek with Prigogine: Is Žižek a Philosopher of Complexity? (Unpublished master’s thesis). Bilgi University, Istanbul, Turkey.

Erdin, H.O. (2006). *Berry’s Phase* (Unpublished master’s thesis). University of Texas at Austin, USA.

Pre-prints

Erdin, H. O. (2010). Pattern Equivariant Representation Variety of Tiling Spaces for Any Group G. *arXiv:1002.1460v1 [math.GN]*.

ACKNOWLEDGMENTS

I would like to thank Lucas Thorpe, Mark H. Bickhard, and Kenneth R. Westphal for their advice and support during my PhD study. I would also like to thank Boğaziçi BAP-D grant 18B02D1 for partially funding my PhD thesis.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

Attempts to apply the mathematical tools of dynamical systems theory to cognition in a systematic way has been well under way since at least the early 90s and has been recognized as a serious "third contender" to computationalist and connectionist approaches (Eliasmith, 1996). Nevertheless, it was also realized that such an application will not lead to a robust paradigm as straightforwardly as was initially hoped (Eliasmith, 1996; van Leeuwen 2005). This thesis addresses methodological problems which result from two features of dynamical systems theory: First, dynamical systems theory is a mathematical field; when used in a non-mathematical field (such as physics, cognition, economics, etc.), it does not directly prescribe an ontology or interpretation. Second, dynamical systems theory has sufficient expressive power to model almost anything regarding changing quantities. These two features generate at least two methodological problems. One results from underdetermination of theory by data: for any given target phenomenon, without an initial restricting paradigm, one can easily devise infinitely many different kinds of models using the tools of dynamical systems theory. Second, even if research groups try to devise "local" paradigms for their models, relations between the models developed by different groups cannot be established by dynamical systems alone. Consequently, a patched-up research field emerges. Various forms of ad hoc-ness underlie these issues; following Lakatos (1970) four distinct forms are identified and examined in this thesis:

- Ad hoc$_1$-ness: A new theory is ad hoc$_1$ if it does not have excess theoretical content over its competitors (§3.1).

- Ad hoc$_2$-ness: A new theory is ad hoc$_2$ if none of its excess content is corroborated (§3.1).
- Ad hoc$_3$-ness: If a series of theories are generated by a heuristic that does not provide "continuity" or "glue" between the theories, then the heuristic is ad hoc$_3$. Heuristics which are not ad hoc$_3$ promote the series of theories into a "research programme" (§3.2).
- Ad hoc$_4$-ness: If a series of theories are generated by a heuristic which is not "autonomous" then the heuristic is ad hoc$_4$ (§3.3).

For at least three reasons such ad hoc-ness is an issue. One is local and internal: rather than obtaining a target phenomenon and its explanation, we end up with a target phenomenon together with an infinite set of models with no principled criteria to select between them. The second reason is comparative: there are already research frameworks that do not suffer from these issues (at least not to the same degree), namely, the computational theory of mind and interactivism. This makes the decision to pursue a strict dynamicist approach highly non-trivial. The third reason is global, namely, the present author believes in a version of non-reductive naturalism that posits as a regulative ideal the elimination of all ad hoc-ness (Bickhard, 2020). A consequence of ad hoc-ness related to having merely local models is that although one obtains a more precise mathematical description of the phenomena, it is only a description, not a genuine explanation. An associated danger here is the emergence of circularities which may go unnoticed. For example, blurring the distinction between assuming that representations are not necessary, or instead having this follow consequently (van Leeuwen, 2005).

These issues confronting the pure dynamicist approach have been noticed and various proposals have been made to address them. These proposals range from directly positing ontological instantiations of dynamical systems (van Gelder, 1999) to treating the use of dynamicist tools as metaphor (Thelen & Smith, 1994) to attempts at transcending the metaphorical (Eliasmith, 2003) – who also aims to move further from the "mind as computer" metaphor. This thesis does not offer another dynamicist proposal; instead, I explicate a method for assessing such proposals by drawing upon Lakatos's (1970) *methodology of scientific research programs* (hereafter: "MSRP"). MSRP focuses on the heuristics of a particular field and gauges the model/theory building stratagems by reference to theoretical and empirical progress, on the one hand, and the *continuity* and the *autonomy* of the field's heuristic, on the other. These factors afford distinct senses of ad hoc-ness, which serve as an effective tool to detect various subtleties which may otherwise be missed.

For example, the present approach identifies shortcomings missed by Chemero's (2009) *radical embodied cognitive science*. After noting the shortcomings of a pure dynamicist program, Chemero develops his positive alternative by first identifying a set of norms any mature science should satisfy and then offering two methodologies, one model-based and one theory-based, which satisfy these norms. As we will see in §2, his norms for a strong scientific methodology are having a guide to discovery, being as extensive as possible, and being realist.

MSRP's appraisal, in contrast, reveals additional methodological levels that Chemero's (2009) norms miss. I shall argue that the MSRP appraisal shows that neither his model-based approach nor theory-based approach is as strong as the computationalist or the interactivist methodology (but nevertheless better than the pure dynamicist methodology). Thus, Chemero's radical embodied cognitive science cannot be methodologically as strong as he claims it to be.

The most salient and original contributions in this thesis are the following. First and most important, MSRP must be incorporated into the current debates in the philosophy of cognitive science regarding methodology and explanation. In particular, the notions of continuity and autonomy must be included in the toolbox of any methodological appraisal in cognitive science. This is because, besides their use in appraisals proper, they help identify two tendencies in such debates. One tendency is treating any modelling with differential equations as methodologically equivalent to any other model. Instead, one must focus on the type of the heuristic that constructs models. This makes it clear, for example, the difference between a Newtonian modelling and pure data fitting (§5.1). The second tendency is blurring the distinction between “extensions” of a model and its “applications” (§5.2). The continuity requirement of MSRP blocks such a conflation. One other original contribution in this thesis is an attempt at incorporating control theoretic system organization analysis of the notion of autonomy given by interactivism into MSRP (§7.2). Given the importance of the notion of autonomy for appraisals and given its ambiguous formulation in MSRP, such an attempt, if successful, not only provides further distinctions that are not available within MSRP but also provides additional structure to the types of processes within a heuristic.

Regarding the structure of the thesis, the next chapter will begin by introducing Chemero's arguments for anti-representationalism (§2.1) and his critique of pure dynamical systems modelling of cognitive phenomena (§2.2). He compares the contemporary debate regarding the methodological issues of anti-representationalism to the debate between the thermodynamicists of the 19th century. Noticing similarities, he then suggests dynamicists should adopt similar norms of the "winning" side, especially emphasizing the "guide to discovery" property of the atomists' heuristic. Using these new norms, he argues that anti-representationalist cognitive science can methodologically be on a par with, say, computationalism, if it aims for extensive *models* that provide guide to discovery (the model-based remedy) or uses Gibson's ecological psychology as a background realist *theory* (the theory-based remedy). These solutions and his arguments are described in subsections 2.3.1 and 2.3.2, respectively.

Chapter 3 introduces Lakatos's MSRP (Lakatos 1970, 1976). Lakatos's key point is that a scientific research field must have a metaphysical *hard core* (a *negative heuristic*) which cannot be refuted by facts, and a *positive heuristic* that provides the principles in accord with which practitioners can build *protective belts* of auxiliary hypothesis around this core against various challenges. To clarify important features of MSRP, I draw upon Clark's (1976) appraisal of the methodologies of atomists and phenomenological thermodynamicists in the 19^{th} and early 20^{th} century. Although Chemero (2009) also uses Clark's article, he does not utilize its full potential. Sections 3.1 to 3.3 trace the norms required for a mature scientific field beginning with Lakatos's requirement of growth and ending with the requirement of *autonomous continuous growth*. During this progression 4 types of ad hoc-ness are identified. This thesis is focusing exclusively on the last two types of ad hoc-ness, namely, whether the models and theories generated within a research field form a "continuous" series and whether these models and theories are generated in an "autonomous" manner (sections 3.2 and 3.3, respectively). The strongest methodologies satisfy the requirement of autonomous continuous growth, namely, if it is neither ad hoc_3 (i.e., a patched-up research field) nor ad hoc_4 (i.e., not autonomous).

Chapter 4 utilizes these norms to give appraisals of the pure dynamical approach and computationalism. This is done by explicitly giving the heuristics of these frameworks. The appraisals show that the pure dynamical stance is both ad hoc_3 and ad hoc_4 while computationalism is neither. Although this agrees with the conclusion of Chemero's critique of these methodologies, it reaches that conclusion in a different manner.

In the next chapter, the critique of radical embodied cognitive science proper begins. In §5.1, I argue that Chemero has a reductive interpretation of the 19th century debate as delineated by Clark (1976). Missing the "deeper" lessons there, he ends up settling on suboptimal norms which then causes his remedies to be not as strong as he claims them to be. The appraisal of his model-based approach is given in §5.2 and his theory-based approach is given in §5.3. These sections show that radical embodied cognitive science does not have a methodology as strong as computationalism, so that Chemero's claims to the contrary are false. Finally, since one of the motivating norms for Chemero for extending his model-based remedy to a theory-based remedy is realism, I present some remarks regarding MSRP and realism in §5.4.

Interactivism is introduced and its appraisal is given in sections 6.1 and 6.2, respectively. The importance of interactivism for this thesis is threefold. First, it is an action based but representationalist paradigm of cognition. Thus, it shows that the usual arguments by elimination for anti-representationalism (including Chemero's) are unsound. Second, its appraisal shows that it is methodologically as strong as computationalism and because it subsumes computationalist type representations as a special case (§6.1), it is probably stronger. Third, interactivism's understanding of autonomy can be valuable for making MSRP's notion of autonomy more precise.

§6.3 compares some methodological maneuvers of Gibson and Chemero when confronted with the conceptual issues of correspondence type representations.

Chapter 7 has two parts. §7.1 delineates the status of MSRP with respect to such issues as mere description *versus* explanation and covering-law *versus* mechanistic accounts of explanation. §7.2 is a series of remarks regarding whether interactivism's utilization of far from equilibrium thermodynamic systems for modelling autonomy can be used for MSRP itself.

Finally, Chapter 8 concludes the thesis with open questions and further possible research directions. It also emphasizes Lakatos's insistence that appraisals of MSRP must not be taken as heuristic advice in deciding which research field is the "best."

CHAPTER 2

CHEMERO'S VISION FOR EMBODIED COGNITIVE SCIENCE

In this chapter I will closely follow Chemero's line of reasoning and see why he thinks the pure dynamical view has methodological problems (§2.1 and 2.2). I will then review his two proposals to remedy this situation (§2.3). His first proposal will be to offer a purely dynamical model together with its extensions as a research methodology which has a guide to discovery. Nevertheless, he thinks this approach lacks realism and doubt whether it can really be extensive enough regarding all aspects of cognition. His second, major proposal, will be to appeal to Gibson's ecological psychology as a realist guide to discovery which is extensive enough.

2.1 Chemero's argument for anti-representationalism

Representational theory of mind gradually became the dominant view in philosophy of cognitive science especially after Block and Fodor’s arguments that thought is *productive* and *systematic*, and that representations with processes on them is the best way to capture these aspects (Block & Fodor, 1972; Fodor 1975).[1] During the 80s and the 90s two kinds of further developments occurred. First, the types of representations and their role in cognition diversified. As we will see in this section one important variant is action-oriented representations. Second, with the advent of connectionist and dynamicist views of mind some began to question whether we really need the notion of representation in cognitive science. These anti-representationalists range from claiming that there are no representations in cognition to being agnostic about such a claim. These latter maintain that we do not need

[1] See §4.2 for more on Fodor’s theory.

representations in our explanations of cognitive phenomena. Chemero belongs to this camp. In this section, I will first review Chemero's definition of representation and reconstruct his argument against representationalism. I will then argue that even if one accepts his argument as valid, it is nevertheless unsound.

Chemero (2009) bases his understanding of representations on Millikan's teleological theory (Millikan, 1984, 1993) and gives an if and only if definition as follows:

> A feature R_0 of a system S is a *Representation for S* if and only if:
> (R1) R_0 stands between a representation producer P and a representation consumer C that have been standardized to fit one another.
> (R2) R_0 has as its function to adapt the representation consumer C to some aspect A_0 of the environment, in particular by leading S to behave appropriately with respect to A_0, even when A_0 is not the case.
> (R3) There are (in addition to R_0) transformations of R_0, R_1 . . . R_n, that have as their function to adapt the representation consumer C to corresponding transformations of A_0, A_1 . . .A_n. (pp. 50-51)

For him some of the important properties that the above definition bestows upon representations are being teleological with normativity built in (function in R2 is teleological), requiring representations to be part of a system so that "nothing can represent just one (token) environmental situation" (R3), and that the definition covers *action-oriented*[2] variety of representations (p. 52).

Chemero (2009) then compares this definition to newer definitions of representation and argues that it is just as good as those if not better. He considers three types of candidates for representation that roughly correspond to *effective tracking*, *noneffective tracking*, and *registration* (p. 65). In effective tracking, a part of the organization is in continuous contact with the relevant part of the environment

[2] Action-oriented representations are defined as "representations that both describe a situation and suggest an appropriate reaction to it; they are essentially representations of affordances" (Chemero, 2009, p. 187). Chemero (2009) uses it as an umbrella term that includes *indexical-functional*, *pushmi-pullyu* or *emulator representations* (p. 26).

and the ensuing contentful internal states can function properly only if the contact is continuous. Chemero says that he will “refer to any theory of representation that calls all contentful internal states representations (even if they are not decouplable from their targets) *traditional theories of representation*” (p. 54). According to him this includes the theories of Dretske (1981), Millikan (1984), and Fodor (1990) (Chemero, 2009, p. 65) – as well as his definition of representation above. In noneffective tracking, the contact with the environment can have intermittent disruptions. Thus, the internal states, to be representations, must fulfill the condition of being a temporary “stand-in” in such systems (p. 55). According to Chemero, Grush’s (1997) *emulator theory of representations* is of this variety. According to this theory there are emulators within the organism that predicts the “next state” given the current state. The paradigmatic cases are ballistic movements such as moving an arm and hand toward a target (Chemero, 2009, p. 61). Chemero critiques this view by arguing that stand-ins in the emulators are not that different from the representations of the traditional theories and that since emulators only represent “parts of an animal’s body and not parts of the external environment” (p. 64), it is better to *not* ground the notion of representation on them.

Finally, *strong decoupling* yields the third variety of representations. According to this view the only internal states worthy of having representational properties are the ones that can keep track of their target even when the target is not present for extended periods (e.g., months) (Chemero, 2009, p. 57). This is called “registration” (Smith, 1996). An agent *registers* an object via detachment and abstraction from detail. Note that this also requires reidentification. Chemero thinks that grounding the notion of representation on registration is too restrictive (Chemero, 2009, p. 59). Moreover, he does not want to offer anti-representational arguments that are based on merely changing definitions (p. 66). Thus, he prefers the traditional theories of presentations as the target of his critiques which we not turn.

Accepting the Millikan type definition of representations given above, Chemero examines their ontological and epistemological status in cognitive science. For example, an anti-representationalist might claim that they do not exist in cognition because animals are like any dynamical system, but only more complicated. Therefore, since dynamical systems lack representations, so do animals. Chemero (2009) calls this the "metaphysical claim" (p. 67). As for the epistemological status of representations, one can claim that "our best explanations of cognitive systems will not involve representations" and he calls this the "epistemological claim" (p. 67). Note that if the metaphysical claim is true, then it must be the case that our best explanations of cognitive systems will not involve representations at the most fundamental level. Conversely, it can be argued that if our best explanations do not posit representations then it is natural to claim that there are no representations.

Now, an anti-representationalist can defend various combinations of these non-existence claims (Chemero, 2009, p. 68). That is, one can claim that there are no representations, and our best explanations will not use them, but also that even if there are no representations, it is best to use them in our theories. This can be because they might provide other benefits such as model building, render calculations tractable, or point to novel experiments. In such a case, representations are ascribed only *instrumentally* (p. 67).

What is interesting for Chemero (2009) is that one can deny or simply be agnostic about the metaphysical claim but defend the epistemological claim. That is, one cannot know whether representations exist, but it is best not to use them in our theories. He argues for this last position by first analyzing the Watt governor, the paradigmatic counterexample to arguments for the necessity of representations. Chemero claims that it can indeed be seen as having representations! If this is so, then the metaphysical claim loses one of its major supports.

Watt governor is a mechanical device that stabilizes how fast a steam engine turns its shaft. The rotation of the shaft is correlated with the rotation of a spindle in such a way that as the engine produces faster rotations, the arms attached to the spindle go higher because of higher centrifugal forces. This in turn causes a valve to restrict the flow of steam which, in turn, reduces the rotation rate of the shaft. Conversely, if the rate drops too much, then the arms of the spindle also rotate in a slower rate causing an inward motion of the arms and hence a lower rotation angle. This in turn opens the valve so that the ensuing increased rate of steam flow causes the engine to operate faster. By carefully adjusting the weights in the spindle arms and the correlations between the engine, valve, spindle turning rate, and spindle arm angle, Watt was able to achieve sufficiently stable turning rates for certain industrial applications.

Now, the details of how Chemero ascribes representations to Watt governor does not matter for the purposes of this thesis, but 3 remarks are in order:

I. The valve of the Watt governor is the representation consumer, and the spindle is the representation producer. The angle of the arm is the action-oriented representation since it describes both the situation and the needed action (i.e., the speed of the engine and the amount of change needed in the valve opening) (Chemero, 2009, p. 71). Note that many dynamicists will deny such an ascription of representations to the Watt governor. Indeed, van Gelder (1995) introduced the Watt governor to the philosophy of cognitive science community precisely to demonstrate that a system without representations and computations can accomplish complicated tasks.

II. Although representational, it is nevertheless *not* computational because there are no inferences made by the system using these representations (Chemero, 2009, p. 72).

III. One worry for me is that because the Watt governor is such a classical Newtonian mechanical system, there is a danger of over generalizing the claim that it has representations to *any* dynamical system. For example, what prevents one to claim that the solar system also has action-oriented representations? Chemero is aware of such a possibility (Chemero, 2009, p.74) but he nevertheless does not discuss its more general consequences, say, for the whole of physics.

According to Chemero (2009), this situation of the Watt governor enables one to see that the epistemological claim might be true. The explanation of the Watt governor in terms of differential equations are way more aesthetic, powerful, and economical compared to any explanation using representations. The takeaway for Chemero is what he calls the "dynamical stance" (p. 72). This means that one should be agnostic regarding the status of the metaphysical claim and stick to the epistemological claim together with the tools of dynamical systems for anti-representational modeling.

Now, even if one accepts Chemero's argument as valid, it is nevertheless unsound. He implicitly assumes that the criteria he proposes for what a representation is are the only possible ones. Note the form of his argument: State the (only possible) definition of what a representation is. Then, show that it renders the metaphysical claim indefensible and that it supports the epistemological claim. Since the only possible (type of) definition of representation causes such problems, it is better to have anti-representational theories. But, if there are other models of representation which do not succumb to the same problems, then the above conclusion does not follow (Bickhard, 1993, 1997, 1998a).[3] I agree with Chemero that representations as given by his definition have grave issues[4] but do not agree with his conclusion, namely, that anti-representationalist theories are thus better.

Although Chemero is an anti-representationalist (and a radical one), he nevertheless does not find the dynamical stance as presented above satisfactory. In the next section I will review his norms for what it takes to be a good methodology of a research field and his critique of the dynamical stance.

2.2 Chemero's philosophy of science: A lesson from physics

So far, we have seen why Chemero thinks one should be an anti-representationalist: Since representations can be applied to almost any dynamical system, they lose their explanatory value. Moreover, these explanations are usually clumsy compared to the elegance of differential equations. Hence it is better to accept the dynamical stance, which is agnostic when it comes to the metaphysical claim (i.e., there really are no representations in cognition), and endorses the epistemological claim (i.e., our best explanations of cognition do not need to posit representations). In this section I will present Chemero's arguments for why such a dynamical stance has methodological problems and then delineate his 3 norms for a good enough scientific methodology. These will be providing a *guide to discovery*, being as extensive as possible, and telling what cognition really is, namely, *realism*.

---

[3] This kind of critique has been applied to Fodor and even to Gibson (see §6.3 below; see also Bickhard & Richie, 1983; Bickhard & Terveen, 1995).

[4] But not necessarily the same issues he considers.

According to Chemero (2009) the methodological problem with the dynamical stance can be discerned as part of the general debate between the computationalists and the dynamicists. Unlike computationalists, dynamicists seem only to have some ad hoc applications of differential equations and data fitting techniques. More importantly, computationalists have a framework that lets them make predictions and design experiments. The predictions of the radical dynamicists are restricted to the particular cases for which the equations were tailored. To clarify his position regarding the shortcomings of the dynamical stance, Chemero compares this debate to a similar one between physicists in the late 20th and early 21st century (p. 78).

This debate was between the *atomists*, represented by Boltzmann, and the *phenomenalists,* represented by Mach.[5] Mach was against the idea of positing non-observable entities. In particular, he was against positing atoms to explain various observed phenomena in nature. Note that even in the second half of the 19th century, there still were not any arguments or experiments that managed to convince the scientific community regarding the existence of atoms. Against Mach, Boltzmann presented an argument that Chemero (2009) calls "the guide to discovery argument" (p. 79). Boltzmann argued by first asking whether "the atomistic theory has had many remarkable specimens to show, even long after the period of its greatest glory"[6] (Boltzmann, 1900, p. 253 - in Chemero, 2009, p. 79). He argued that all the recent success of physics in explaining such phenomena as chemical substances via Van der Waals' formulas or hydrodynamics were not possible without the atomistic

[5] In §3.3 we will see the nature of this debate more extensively covering many other physicists besides Boltzmann and Mach via following Clark's paper (Clark, 1976). Although Chemero cites Clark's paper he does not use the full extent and power of it.
[6] As we will see in §3.3, Boltzmann has the *degenerating phase* of the atomistic research programme towards the end of the nineteenth century in mind when he states that its glory days are over.

assumptions. Moreover, in contrast to the phenomenologist methodology, atomism "enables us to point definitely to the experiments which are in most likelihood to lead to its correction" (Boltzmann, 1900, p. 254 - in Chemero, 2009, p. 79). Thus, atomism gives us a guide to discovery for coming up with new experiments and with new laws. As Chemero (2009) says, phenomenological physics can only be improved by "ad hoc additions to the theory, in light of new empirical facts. That is, phenomenological physics, because it refuses to postulate underlying, unobservable structure, provides no guide to discovery" (p. 80). This *fact dependency* renders phenomenological physics incapable of offering any guide to discovery.

Returning to the current debate in cognitive science, Chemero (2009) then claims that the computational representational theory of mind is like atomism, a theory with guide to discovery, whereas an embodied approach adhering only to dynamical stance is not (p. 81). Thus, Chemero argues that the "question, then, is whether dynamical cognitive science can avoid being purely fact dependent, despite the fact that it posits no underlying mechanism" (p. 81). Because of the above considerations, his initial answer is a no. Then, he suggests two possible solutions. First, that "dynamical models themselves can be guides to discovery" and second, that "radical embodied cognitive scientists can use Gibsonian ecological psychology as a guide to discovery just as physicists of the early twentieth century used atomism, that is, as a non-instrumentalist background theory of the nature of the subject matter"[7] (p. 83). In the next section I will review these two offers for a better research methodology.

[7] There seems to be an ambiguity in this sentence if it is taken as an answer to the above quoted question. On the one hand, Chemero can be taken as claiming that Gibsonian psychology can be utilized as a guide to discovery without positing an underlying mechanism and this will nevertheless be as good as the atomists' guide to discovery in the 19th century. On the other hand, a stronger reading of it would claim that Gibsonian psychology can itself be utilized as providing an underlying mechanism so that the situation will be exactly like the atomists of the 19th century where Newton's

2.3 Chemero's solutions to the problem: *HKB+* and Gibson

Chemero (2009) offers one minor and one major solution for radical embodied cognitive science to have a guide to discovery. The minor one is that dynamical models themselves can be guides to discovery, say, by using an extensible model like the one developed by Haken, Kelso, and Bunz (1985) for their finger wagging experiments (hereafter: "HKB"). Chemero uses this model and its extensions (hereafter: "HKB+") as a case study for his claim that "the lack of posited underlying mechanism need not be a problem for radical embodied cognitive science" (p. 95). Then, he presents a critique of HKB+ to motivate his major Gibsonian solution. I will begin with HKB+.

2.3.1 The HKB+ case

The aim of the HKB was to determine the extent to which “limbs in coordinated actions could be understood as non-linearly coupled oscillators whose coupling requires energy to maintain” (Chemero, 2009, p. 86). Subjects were asked to move both of their index fingers horizontally, slowly increasing their speed. It turned out that there were only two stable patterns. One was called “in-phase” where the fingers oscillate in such a way that they keep simultaneously approaching and then diverging from each other. The other was called “out-of-phase” where both fingers oscillate in parallel, horizontally (Chemero, 2009, p. 86). The subjects were first asked to begin wagging their fingers slowly in the out-of-phase mode and then slowly increase their wagging while trying to keep this pattern. But as the speed increased, they were unable to keep it and end up switching to the in-phase mode (Chemero, 2009, p. 87).

laws of motion were the underlying mechanism. His position, as eventually becomes clearer, is the first.

To model this behaviour with differential equations, the relative phase ϕ between the fingers were chosen to be the time dependent parameter and a differential equation describing the behaviour of ϕ was obtained by considering the "simplest possible potential function that will capture all of the data on the finger-wagging" (Chemero, 2009, p. 87).[8]

By investigating further properties of this equation, HKB showed that the fit with the data was remarkable (Chemero, 2009, p. 88). For example, the equation predicts that there should be a critical value where $\phi(t)$ begins to fluctuate erratically before settling in one of the two stable solutions. This behaviour was indeed observed in subjects' finger motions. Such a robust match between the predictions of the equation and the behaviour of the subjects naturally raises the prospects of whether all behaviour can be modelled like this. Admittedly, this particular task was not very *representation hungry,* but surprisingly, it turned out that an HKB-like model can also be applied to such problems (Chemero, 2009, p. 93). In *gear problems*, for example, a picture of connected gears are shown to subjects and the task is to figure out which way the target gear turns as a result of the rotation of the source gear. While the subjects are working on the problem, their finger movements are observed: The moment they figure out the easy way to solve the task is always preceded by an increase in fluctuation in their finger movements. Here, the easy way to solve the problem is to note that the odd numbered gears will turn in the same direction as the source gear. Hence, like the original HKB case, we have two stable loci (that is, the hard way to solve the problem and the easy way) and the cognitive system switches from one to another after fluctuations (Chemero, 2009, p. 94).

[8] The equation was $d\phi/dt = -A \sin(\phi)-2B \sin(2\phi)$, where A and B are called the "parameters" of the equation. The ratio B/A is called the "control parameter" and depends on the frequency of the wagging (Chemero, 2009, p. 88).

According to the computationalist approach this “a-ha” moment is thought to be the learning of a new representation (Chemero, 2009, p. 95). However, it can also be modelled without representations as a shift from one stable solution to another of the governing differential equation depending on the value of the parameter in the equation.

In Chemero (2009) there are six other modifications or extensions of the HKB in which the models' predictions were confirmed (pp. 86-96). According to him this shows that HKB together with its extensions satisfies the requirement of guiding discovery and, to an extent, also the requirement of “being as extensive as possible.” This in turn means that such a “generalizable dynamical model” can promote pure dynamical cognitive science to a research field which is dynamical, anti-representational, and has a methodology that is on a par with the methodology of computational and representational cognitive science (p. 85). Moreover, HKB offers its own successor in terms of a "general science of coordination dynamics" (Chemero, 2009, p. 96)[9] which being rooted in far from equilibrium dissipative coordination of coupled oscillators has the prospects of a "claim about what coordinative structures *are*" (Chemero, 2009, p. 96). He nevertheless says that this is not specific enough to be a realist approach and thus one actually ends up having an instrumental model in practice: "So the significant advances to our understanding that have been brought about by the HKB model are not purely instrumental in origin, even if they are in practice" (Chemero, 2009, p. 96). Accepting these observations, we can see that for Chemero (2009) even though a model can have a guide to discovery it might still not be good enough. In addition to realists being not satisfied with the HKB, he says that since "a good deal of cognition seems not to be

[9] As we will see in §3.2., the ability to generate its own successors in the *right* way is a very important feature of *good* research programmes.

coordinated activity" (p. 96), one needs a more extensive framework. Thus, according to Chemero, for a candidate to be a *theory*, being a guide to discovery must be supplemented with being as extensive as possible and realist, namely, "tells us what cognition is - a push, that is, toward a *theory* of cognition" (p. 97).

Thus, Chemero now has the full set of criteria that will rescue radical dynamical cognitive science from having an inferior methodology. The key to a solution that will satisfy these criteria, Chemero argues, is to be found in the Gibsonian ecological psychology.

2.3.2 The Gibsonian case

Chemero (2009) claims that the three basic principles of Gibson can ground a guide to discovery for radical embodied cognitive science (p. 98). These three principles are the claims that

I. perception is direct,
II. perception is for action, and
III. perception is of *affordances*.

He gives a cursory explanation of these three principles as follows (postponing their detailed accounts to later two chapters). Perception is direct means that it is not "the result of mental gymnastics, of inferences performed on sensory representations" (p. 98). He states that "this implies, of course, that the perceiving isn't inside the animal, but rather is part of a system that includes both the animal and the perceived object" (p. 98).

For the second principle he says that the "purpose of perception is for the generation and control of action. It is usually added to this that a good deal of action is also for perception or cognition" (Chemero, 2009, p. 98). This is Gibson's familiar insight that we act by moving our eyes, head, and body to generate perception and the results of this perception are primarily for action. Note that this principle is compatible with both kinds of embodied cognitive science; radical or otherwise.

The last principle, according to Chemero (2009), follows from the first two. Basically, the environment and the organism's direct perception for action should be rich enough in information to guide the organism's actions. This richness is provided by what Gibson called "affordances" which Chemero provisionally defines as "directly perceivable opportunities for action" (p. 99).

Chemero’s (2009) aim at this stage is to suggest "that a theory based on these three principles can serve as a guide to discovery for radical embodied cognitive science, providing a background theory about what cognition is that can be part of a successful scientific endeavor" (p. 99). He motivates this claim as follows. First, models are necessary for any proper science since "theories are far too complex to be tested against empirical phenomena" (p. 99). One chief reason for this is the computational intractability of sticking only to fundamental entities of the background theory.[10] For example, in biology considering only the DNA and various other molecules will render it impossible to deal with, say, respiratory systems of mammals. Even if the system is tractable enough where there are explicit equations available, it might still be the case that there might not be tractable solutions to those equations. A classic example for this is the Newtonian gravitation equations for 3 masses such as the Earth-Sun-Moon system. Chemero states that models are mediators between theories and data, and that "the appropriateness of a theory as a guide to discovery, then, is partly a function of how well it does at generating models

[10] As we will see in §3.3, there is a stronger reason for the necessity of models, namely, the *metaphysical core* of a research programme cannot be directly tested or refuted by any empirical fact in principle.

for application to laboratory findings" (p. 100). Chemero claims that Gibsonian ecology is clearly a source of inspiration for such models, and indeed that HKB was initially inspired by Gibsonian considerations.

As for realism, in the last chapter of his book, Chemero defends a kind of realism based *entity realism* which he calls "phenomenological realism" (Chemero, 2009, p. 198). The details of this kind of realism is beyond the scope of this thesis but Chemero (2009) argues that affordances are "part of the basic furniture of the universe" (p. 193).[11]

Having now reviewed Chemero's norms for a good enough scientific methodology, in the next chapter I will delineate Lakatos's norms. We will see that realism need not be one of the basic criteria. The *way* the methodology generates models will take center stage, and it will turn out that fact dependence together with guide to discovery should be treated in a more subtle way than Chemero's approach.

[11] But see §5.4 for some remarks regarding realism as a strategic move.

CHAPTER 3

MSRP: THE REQUIREMENT OF *AUTONOMOUS CONTINUOUS GROWTH*

In this chapter Lakatos's (1970) *methodology of scientific research programmes* (hereafter: "MSRP") will be introduced. Each of the modifiers in "autonomous continuous growth" have particular senses in MSRP so I will build them up one by one in what follows.[12] Clark's (1976) contributions to MSRP are crucial both for Chemero and my approach, so a detailed account of his analysis of the methodological issues in 19th century physics will be presented also.

Corresponding to a lack in the "right" type of growth, continuity, and autonomy the following types of ad hoc-ness are considered:

- Ad hoc$_1$-ness: A new theory is ad hoc$_1$ if it does not have excess theoretical content over its competitors (§3.1).
- Ad hoc$_2$-ness: A new theory is ad hoc$_2$ if none of its excess content is corroborated (§3.1).
- Ad hoc$_3$-ness: If a series of theories are generated by a heuristic that does not provide "continuity" or "glue" between the theories, then the heuristic is ad hoc$_3$. Heuristics which are not ad hoc$_3$ promote the series of theories into a "research programme" (§3.2).
- Ad hoc$_4$-ness: If a series of theories are generated by a heuristic which is not "autonomous" then the heuristic is ad hoc$_4$ (§3.3).

[12] Lakatos (1970) does use the "requirement of continuous growth" as one of his titles (p. 173). Also note that in this thesis I will not pursue the broader issues of Lakatos's theory within the philosophy of science such as its status with respect to Popper's or Kuhn's theories other than a few scattered remarks.

3.1 The requirement of *growth*

According to Lakatos (1970) the unit of appraisal for a research field is not an isolated theory, but a series of theories (pp. 117-120). One reason for this is that any theory and a set of factual propositions can be made compatible by adding suitable auxiliary hypotheses. Hence it is important to identify norms that distinguish which harmonizing adjustments are admissible and which are not. Those that are not are ad hoc.[13]

Lakatos (1970) calls a series of theories "theoretically progressive" or "constitutes a theoretically progressive problem-shift," if each member of the series has excess empirical content because it predicts an unexpected fact (p. 118).[14] It is called "empirically progressive" or "constitutes an empirically progressive problem-shift," if some of this excess content is corroborated. If it is both theoretically and empirically progressive, then it is simply called "progressive." Finally, if it is not progressive, it is called "degenerating."[15] For example, if someone can come up with a Cartesian explanation of how Newtonian gravitation, which acts at a distance, results from mechanical contact forces, then the appraisal of whether this is a progressive or degenerating step depends not on how elegant or ingenious the explanation itself is but on whether this new theory in the series predicts some further novel fact (p. 126).

These distinctions allow Lakatos (1970) to make corresponding distinctions among senses of ad hoc-ness. He calls new theories with no greater scope than their

---

[13] Four different senses of ad hoc-ness will be discussed in what follows.
[14] The sense of "novel fact" that I endorse takes into consideration the way a research programme is constructed. This implies that novelty is not restricted to temporal novelty: a well-known empirical fact might be novel for a research programme if it was not taken into consideration during its construction (Zahar, 1976, pp. 216-18). This sense of "novel fact" was later endorsed by Lakatos (Lakatos & Zahar, 1978, pp. 184-89). A famous example is the prediction of the precession of Mercury's orbit via general relativity. Although it was already known by the mid-19th century, it is nevertheless a novel fact with respect to general relativity.
[15] Note that empirical progress implies theoretical progress.

predecessors "ad hoc_1" (p. 175, fn. 2). Theories that predict novel facts, though none are corroborated, are called "ad hoc_2."[16] These distinctions also hold of series of theories: Any series of theories is either ad hoc_1 or is theoretically progressive; any series of theories is either ad hoc_2 or is empirically progressive. However, even a series satisfying both norms may yet confront the "tacking paradox" (p. 131). According to this objection, a low-level hypothesis completely disconnected from the original theory might be added to it to solve a particular problem, so that the resulting series can be progressive (in the sense defined above). Instead, constituents of the new hypothesis must be relevant to the theory they supplement; they must be "*more intimately*" related than mere logical conjunction. Therefore, a heuristic uniting a series is important. According to MSRP, a heuristic which unites a series in a *continuous* manner transforms that series into a *research programme*.

3.2 The requirement of *continuous* growth

"Continuous" in Lakatos's "requirement of continuous growth" might be understood in two senses. According to one, each new theory in a series must provide growth, indeed persistent growth (*per* §3.1). Or this continuity may require a kind or degree of cohesion between the theories within a series. The core idea of MSRP is the following. According to Lakatos (1970), to understand a particular scientific research area and the rationality of its practitioners, one must identify what he calls its "negative and positive heuristics" (p. 132). A negative heuristic is metaphysical and so cannot be falsified by empirical considerations; in particular, it cannot be falsified by *crucial experiments*.[17] Thus it forms the "hard core" of the research programme.

[16] For an example of ad hoc_2 theorizing Lakatos (1970) gives Planck's radiation formula (p. 175, fn. 3).

[17] Note that this is a *methodological irrefutability* in the sense that it is *decision based* and not *syntactical irrefutability*. The latter are about propositions that have the logical form of "for all-there

In contrast, the *positive heuristic* of the programme forges links to refutable empirical content by generating particular models or theories with the help of *auxiliary hypotheses*; and these form a *protective belt* around the hard core (p. 135). Any anomaly or negative finding will be assigned to this protective belt to preserve the hard core. A typical example of these notions is Newton's planetary theory.[18] The hard core (or negative heuristic) of Newton's research programme is his three laws of dynamics plus the gravitational law. The way Newton constructed a series of expanding models around this core is summarized by Lakatos in these terms:

> Newton first worked out his programme for a planetary system with a fixed point-like sun and one single point-like planet. It was in this model that he derived his inverse square law for Kepler's ellipse. But this model was forbidden by Newton's own third law of dynamics, therefore the model had to be replaced by one in which both sun and planet revolved round [sic] their common centre of gravity. This change was not motivated by any observation (the data did not suggest an 'anomaly' here) but by a theoretical difficulty in developing the programme. Then he worked out the programme for more planets as if there were only heliocentric but no interplanetary forces. Then he worked out the case where the sun and planets were not mass-points but *mass-balls*. Again, for this change he did not *need* the observation of an anomaly; infinite density was forbidden by an (inarticulated) touchstone theory, therefore planets *had* to be extended. This change involved considerable mathematical difficulties, held up Newton's work -and delayed the publication of the *Principia* by more than a decade. Having solved this 'puzzle', he started work on *spinning balls* and their wobbles. Then he admitted interplanetary forces and started work on *perturbations*. At this point he started to look more anxiously at the facts. Many of them were beautifully explained (qualitatively) by this model, many were not. It was then that he started to work on bulging planets, rather than round planets, etc." (p. 135)

This is clearly an example of constructing a protective belt around the hard core of Newton's programme, to avoid anomalies, by using the positive heuristic as guidelines for deriving particular forms of Newton's equations for the problematic cases under consideration. This was achieved by setting up coordinate systems and

---

exits" (Lakatos, 1970, p. 183). Such metaphysical statements are a part of the heuristic and hence more than welcome in MSRP.

[18] Other examples of research programmes in Lakatos's sense are relativity, quantum mechanics, Marxism, and Freudianism (Lakatos, 1978, pp. 4-5).

using differential calculus, where the protective belt of auxiliary hypotheses "included geometrical optics, Newton's theory of atmospheric refraction, and so on" (Lakatos & Zahar, 1978, p. 179).

The use of such a heuristic capable of providing such a cohesion implies that appraising a particular theoretical field requires considering both growth and continuity:

> *My account implies a new criterion of demarcation between 'mature science', consisting of research programmes, and 'immature science', consisting of a mere patched up pattern of trial and error*. For instance, we may have a conjecture, have it refuted and then rescued by an auxiliary hypothesis which is not ad hoc in the senses which we had earlier discussed [that is, ad hoc_1 and ad hoc_2]. It may predict novel facts some of which may even be corroborated. Yet one may achieve such 'progress' with a patched up, arbitrary series of disconnected theories. Good scientists will not find such makeshift progress satisfactory; they may even reject it as not genuinely scientific. They will call such auxiliary hypotheses merely 'formal', 'arbitrary', 'empirical', 'semi-empirical', or even 'ad hoc'. (Lakatos, 1970, p. 175)

Considering both growth and continuity makes explicit the corresponding two types of weakness a theoretical series might display. According to Lakatos (1970), although Marxism or Freudianism are unified and have continuity, they are no longer growing (p. 176). On the other hand, although some "modern social psychology" has some novel predictions, their "patched-up, unimaginative series of … adjustments" have no cohesive core (p. 176).

Corresponding to this continuity requirement is a new kind of ad hoc-ness. A "patched up pattern of trial and error" stratagems which nevertheless achieves a progress via an "arbitrary series of disconnected theories" Lakatos (1970) calls "ad hoc_3" (p. 175, fn. 3). Here we must distinguish two senses of ad hoc_3, depending whether there is a heuristic. If there is no heuristic, either because the level of analysis is too broad or because either none has yet been made explicit or because there are more than one, then this amounts only to a patched-up field with an "arbitrary series of disconnected theories." However, if there is a heuristic, then the construction of the newly generated theory has departed from the programme's heuristic. As Clark (1976) says, the new models were generated "in a way which did not accord with the heuristic of the programme" (p. 84), making for discontinuity.[19] This kind of discontinuity can occur when a previously continuous heuristic is no longer able to accommodate anomalies and so enters a degenerating phase. However, such discontinuity may occur at the outset if a well-articulated heuristic itself generates a patched-up series of theories.

For present purposes, these alternatives are disambiguated thus: First, at the foundation of MSRP Lakatos (1970) is unclear about what is a "heuristic." He says that even "science as a whole can be regarded as a huge research programme with Popper's supreme heuristic rule: 'devise conjectures which have more empirical content than their predecessors'" (p. 132). He then says, "what I have primarily in mind is not science as a whole, but rather *particular* research programmes" (p. 132). Unfortunately, the borders of any such "particular" are not specified. Thus, if we have an ad hoc_3 research field, should we say that it lacks a research programme because a heuristic must not be ad hoc_3 to render a series of theories into a cohesive research programme? Or should we simply grant that it *is* an ad hoc_3 research programme? I propose that, if a hard core and a positive heuristic can be identified which are sufficiently "particular," then we should call it a heuristic. This is close to the ordinary meaning of the word "heuristic." Whether such a heuristic is ad hoc_3 is a further issue to be decided. Thus, for the purposes of this thesis, a heuristic which is not ad hoc_3 will promote a series of theories into a research *programme*.[20] Otherwise,

[19] This the reason Lakatos (1970) finds Popper's ban on *content-decreasing* theory generation as too weak (p. 182).
[20] To anticipate, in the next chapter I argue that the pure dynamical approach has an ad hoc_3 heuristic so that according to MSRP it is not a research programme.

that series is ad hoc$_3$ theorizing. The second ambiguity is minor; as Clark (1976) states (quoted just above), ad hoc$_3$ concerns the *type* of a heuristic. However, because MSRP is a dynamic approach, an otherwise sound research programme may become ad hoc$_3$. The important point is that, when such a change occurs, the resulting series of theories becomes patched-up so that, if the series is still counted as one theoretical framework, then it must be demoted from the status of a research *programme*.

Lakatos's (1970) requirement of continuity is how MSRP accounts for unity, simplicity, or beauty in science (p. 131, see also p. 175). However, even if continuity is provided by a well-articulated negative and positive heuristic, there may yet be different kinds of methodological power. This is the "autonomy" of mature research programmes. Although Lakatos offers remarks about how a heuristic provides its practitioners autonomy from "external" input during their investigations, Clark (1976) explicated this aspect of MSRP. Next section reviews Clark's account of *fact dependency* and its relation to autonomy.

3.3 The requirement of *autonomous* continuous growth: A deeper lesson from physics

According to MSRP, Newton's programme is typical of any "mature" research programme: It is constantly surrounded by infinitely many anomalies and puzzles; deciding where to begin with these and how to proceed, in what manner the theories or models should be enhanced, modified, etc., must be relatively independent of *external input*. Internal norms suffice for many key maneuvers and, perhaps more importantly, for assessing the importance or need of such maneuvers. Lakatos (1970) says that such a protective belt "is not built up in an eclectic fashion, without any preconceived order" (p. 135). In other words, the positive heuristic focuses the attention of the scientist on changing or modifying theories by allowing her to ignore "the *actual* counterexamples, the available '*data*'" (p. 135). This capacity to proceed almost automatically without external input, guided by internal ideals, Lakatos calls the "autonomy of theoretical science:"

> *Mature science consists of research programmes in which not only novel facts but, in an important sense, also novel auxiliary theories, are anticipated; mature science — unlike pedestrian trial-and-error — has 'heuristic power'.* Let us remember that in the positive heuristic of a powerful programme there is, right at the start, a general outline of how to build the protective belts: this heuristic power generates *the autonomy of theoretical science.* (p. 137)

Here, *heuristic power* means "the power of a research programme to anticipate theoretically novel facts in its growth" (p. 155, fn. 3).[21] Now, one way to gauge the degree of autonomy a heuristic provides is by Clark's (1976) notion of fact dependence. He distinguished fact-dependent from fact-independent heuristics to understand differences between the research programmes of phenomenological and atomist thermodynamics. According to Clark, the heuristic of phenomenological thermodynamics was fact-dependent even while it was flourishing (before it became a degenerating programme) and this dependence results in a "weak heuristic," compared to the atomist's "strong heuristic." Since Clark's adjectives "weak" and "strong" may be confused with the temporal status of a heuristic, whether it is degenerating or progressing, I will use "ad hoc_4" for Clark's "weak heuristic" (which is strongly fact-dependent) and shall designate his (fact-independent) "strong heuristic" as "not ad hoc_4" heuristic.

As we have seen above (§2.2) the main difference in thermodynamics between phenomenologists and atomists was whether to postulate unobservable entities, such as atoms, to explain thermodynamical phenomena. According to

[21] Note that this includes the capacity to generate "unsolved but solvable problems" (Musgrave, 1976, p. 482).

atomists, any such phenomenon must result from atoms interacting according to Newton's laws. This heuristic provided atomists great autonomy in their derivations because, once some simplifying assumptions are made, the derivations can proceed and predictions can be obtained with little empirical input (Clark, 1976, pp. 47-56). The result must be evaluated against empirical data, but the route to this stage was significantly more autonomous than the phenomenological heuristic.

The way this proceeded was very similar to Lakatos's rendering of Newton's programme (as quoted in §3.2). To see this note that the hard core of atomism is the following: "The behaviour and nature of substances is the *aggregate* of an enormously large number of very small and constantly moving elementary individuals subject to the laws of mechanics" (Clark, 1976, p. 45). The positive heuristic of atomism consists of the following four principles (Clark, 1976, p. 45):

I. Make simplifying assumptions about the specific nature of the atoms and apply the laws of mechanics to their behaviour and interactions.
II. For the aggregate, the individual motions and hence the individual properties can be irregular but there will always exist certain mean values corresponding to those properties given by certain distributions among the atoms.
III. The above two simplifying assumptions should be weakened or even eliminated to better approach the real phenomena under consideration, e.g., a "real" gas as opposed to an "ideal" gas.
IV. Using the specific theory created thus far, its consequences should be investigated to obtain results regarding the internal properties of the phenomena under investigation as well as to derive the known macroscopic properties as limiting cases.

What are the series of theories generated by this heuristic? Clark (1976) identifies four different atomic theories regarding the kinetic theory of gases during the second half of the 19th century (pp. 47-56). The first one is Clausius' elementary kinetic theory which for the first time tried to establish in a systematic manner the connection between mechanical properties of atoms (such as mass, momentum, energy) and the macroscopic properties of gases (diffusion, viscosity and heat, respectively). One of his simplifying assumptions that falls under the heuristic principle 1 above was regarding atoms as smooth elastic spheres. Another one, corresponding to 2 above, was his assumption that "all the molecules can be regarded as traversing rectilinear paths with the same speed, in arrays parallel to and normal to the walls of the containing vessel" (p. 47). As certain anomalies with observed properties of gases emerged, certain adjustments of the model as prescribed by 3 above were made. These include abandoning the assumption of atoms being smooth elastic spheres, abandoning the assumption of perfect elastic collisions, and abandoning the assumption of motion in regular arrays. These occurred in 1857.

Then, in 1860, Maxwell proposed another theory. He began with three assumptions (Clark, 1976, p. 50). Molecules were again to be regarded as elastic spheres but all directions of rebound after collisions were to be considered, and the probability of a component of velocity having a certain value would be independent from its other components. Using his model, Maxwell improved upon the previous models while at the same time staying within the same research programme and its hard core. For our purposes it is not necessary to discuss the specific details of each modification in the next two kinetic theories suffice it to say that the third kinetic theory was again Maxwell's six years after his first one and the major change was that atoms were no longer spheres but centers of forces (Clark, 1976, p. 53). Finally, in 1868, Boltzmann began with Maxwell's theory and got rid of even more of Maxwell's simplifying assumptions (Clark, 1976, p. 57).

As with Newton's case, we see here too that the protective belt is not "built up in an eclectic fashion, without any preconceived order" and the "order is usually decided in the theoretician's cabinet, independently of the *known* anomalies" (Lakatos, 1970, p. 135). This in turn means that such a scientist building increasingly complex models as dictated by the positive heuristic is selective regarding data. Thus, the crucial aspect of the above process for this thesis is the way a mechanistic atomic model can enhance its sophistication independent of *external input*, i.e., in an autonomous manner.

Now the heuristics of phenomenological thermodynamics are as follows. According to Clark (1976) the hard core of the programme is the principle that there is a "*definite relation between a quantity of heat and the work which in any way could be produced by it*" (p. 63). The positive heuristic is as follows (p. 64):

I. Always investigate cyclic processes.

II. All processes should be reversible.

III. Use "particular empirical laws (constitutive relations) to discover … functional relationships persisting among empirically determinable parameters" (p. 64) via 1 and 2 above.

IV. In every conversion of heat into a mechanical effect there will always be some other process that must occur in such a way as to prevent a total conversion of the former to the latter (p. 66). It is important to note that this fourth heuristic was added later to the programme and was in effect an explicit guide to tackle anomalies.

Before seeing why this is ad hoc$_4$, the relevant aspects of the series of theories built according to these principles should be given. Clark (1976) identifies four of them beginning with Carnot's theory in 1824. The remaining three are "the mechanical theory of heat, fully-fledged phenomenological thermodynamics and … the system of the four laws of thermodynamics" (p. 63). For the purposes of this thesis, it is not essential to trace the details of the progression of these theories but since Chemero (2009) contrasts his approach to phenomenological thermodynamics we should delineate certain aspects of the third theory. Clark (1976) discusses Planck's formulation of the two fundamental laws as follows (p. 68):

I. During every process in nature the total energy remains constant.

II. During every process in nature total entropy always increases and never stays constant. Reversible processes are only an idealization.

According to Clark, Planck's move was to generalize Clausius' law, which was specific to "*transfer* of heat and mass *motion*" to all processes (p. 70). Also, note that the second law is universal in the sense that statistical violations of it are not allowed. This was a major conflict with Boltzmann's later theory where the statistical nature of the second law came to prominence in that there can be (local) situations where entropy might *not* increase. But its (global) tendency on the average is always in the direction of increasing entropy.

The gist of the phenomenological heuristic is applying "contentless" laws, namely, the conservation of energy and the non-decreasing of entropy, to particular situations to develop models and theories. In what sense was this heuristic fact dependent, or ad hoc$_4$? Clark captures the inherent problem in a statement by Planck himself:

> From the two fundamental principles … by pure logical reasoning, a large number of new physical and chemical laws [were] deduced, which are

> capable of extensive application, and have hitherto stood the test without exception. (Planck, 1945, p. ix, as cited in Clark, 1976, p. 74)

The fact dependency arises here because the phenomenological heuristic in thermodynamics consists in logically deriving consequences of two extremely general laws by using auxiliary hypotheses derived from the specific empirical problem at hand. Although the atomistic programme also required auxiliary hypotheses, the crucial difference is that the phenomenologists' auxiliary hypotheses were "taken either piecemeal, as low-level empirical generalisations, or are simply borrowed from other theories" (Clark, 1976, p. 75). These hypotheses usually involve relations or formulae between phenomenological variables such as between pressure and temperature and are called "constitutive relations." The content of the two fundamental laws must be derived from "differing situations characterised by different constitutive relations" (Clark, 1976, p. 75). Clark (1976) says the following:

> [Together with the] suggestion as to how to deal with refutations … it is clear that there was no possibility of employing the heuristic to go beyond the two laws, to supersede them. This situation contrasts sharply with the kinetic [i.e., atomist] programme, for there the heuristic laid down quite general guidelines for developing, independently of refutations, new kinetic theories of ever greater empirical content. (p. 75)

Clark states that the heuristic of the atomist programme led to new universal laws (e.g., the non-ideal equations of state) and to increasing empirical content by *systematically* incorporating the fine structure of the heuristic's ontology into this series of theories, so that these new laws were "not mere logical consequences of auxiliary hypotheses and previous kinetic theories" (pp. 75-76). Important here is that this progress is integral to the heuristic, so that there is a high degree of autonomy. Hence this process did not simply follow rules of generic mathematical derivations together with various external inputs whenever they are needed. Because the atomist programme afforded such strong internal progress, anomalies that initially appeared too difficult to accommodate could be postponed to a future time when they may be more tractable. For these reasons, this research programme's heuristic is not ad hoc_4. Note that, as will be important later for us, this includes the independent "determination of the values of crucial parameters from experiment” (p. 76, fn. 152). This is one reason for resisting the lure of “simple” modelling, no matter how brilliant or fast at getting empirical results it turns out to be.

These points may be clarified by a brief, contrasting example of a simple ad hoc_4 heuristic. Clark (1976) claims that the heuristic of the Ptolemaic research programme was ad hoc_4 (although it was so precise as almost to be algorithmic): “*save the phenomena by a combination of uniform periodic motions*” (p. 75, f.n. 149). Here, “progress” reduces to anomaly *dependent* further research, namely, external dependency represented by removing whatever anomaly the scientific community decided was most important. Similarly, the two laws of thermodynamics (conservation of energy and the entropy increase) are high-level universal laws from which all else can be derived by logic and empirical facts, depending upon when and how these facts presented puzzles to be solved. Confronted by each particular external fact, phenomenological thermodynamicists had to use external equations (either empirical or from other theories) to proceed with their derivations. This is the reason their heuristic is ad hoc_4, in contrast to the atomists’.

Let me now summarize these distinctions and then comment on them. Heuristic power is a dynamic property of a temporally extended research programme; a heuristic may degenerate and be replaced by a rival if it is no longer able to account for (*explain*, not merely to accommodate) anomalies or to predict novel facts. Two ways in which a heuristic can be (or become) *ad-hoc* concern theoretical and empirical progressiveness. A heuristic can become *ad-hoc*$_1$ if it can no longer predict novel facts; it can become *ad-hoc*$_2$ if the novel predictions are not corroborated. Recall that Lakatos footnoted a third kind of ad hoc adjustment, which fails to preserve the continuity of a research programme (if there is a *programme* in the first place) or else simply produces a patched-up research field. Such measures are here called ad hoc$_3$. Distinct from these considerations, type of a heuristic can also be gauged by its autonomy: a fact-dependent heuristic counts as ad hoc$_4$. Clark's analysis showed that phenomenological thermodynamics was not ad hoc$_3$ (hence it was a genuine research programme), though it was ad hoc$_4$, whereas the atomist research programme was not ad hoc$_4$. With hindsight we can see that at the end of the 19th century and the beginning of the 20th century, atomism had a strong heuristic but a lack of empirical success whereas thermodynamics had a weak heuristic but strong empirical success. Thus, leading physicists of this era were at a genuine crossroads, and hence it was rational for them to choose any of the programmes to make it better (and there is no in principle reason to not to work in both camps! Clausius was one such example). We will see later that a similar situation exists between the heuristics of the radical embodied cognitive science and representationalism.

These distinctions have three consequences that are important here. First, heuristic power and type of a heuristic are distinct notions, though the heuristic power may bear upon the type of heuristic. If we consider two heuristics, both well into their degenerating phases (marked by the first two kinds of ad hoc adjustments), a heuristic exhibiting significant autonomy would still tackle its anomalies much more autonomously than a heuristic exhibiting much greater dependence upon facts within its domain. We can thus say that a heuristic which is not ad hoc$_4$, has a "core autonomy." Second, by the definition of empirical progress, if a heuristic is not ad hoc$_2$, it is also not ad hoc$_1$. However, for the latter types of *ad-hoc* character (3rd and 4th), Lakatos's view is unclear. Lakatos (1970) may endorse a hierarchy, so that autonomy requires continuity (p. 137). Although this priority seems plausible, whether it holds is unimportant to my further discussion; so, I shall assume they are independent. Distinguishing the *type* of heuristic power from the *degree* of heuristic power affords two distinct types of evaluation of a heuristic. This is important here, as I shall evaluate types of research programme heuristics within cognitive science, rather than their progressiveness.[22]

In the next chapter I use these distinctions of types of heuristic to characterize and to assess computationalism, and the pure dynamical view of mind (Table 1 shows the results of the appraisals of this section).

Table 1. Summary of *Ad-hoc* Types in MSRP and the Results of the Appraisals of Atomism and Phenomenological Thermodynamics

Appraisal at a non-degenerating phase of a given field	ad hoc$_1$ (theory does not also explain other novel facts)	ad hoc$_2$ (none of the novel facts are corroborated)	ad hoc$_3$ (disconnected series)	ad hoc$_4$ (not autonomous)
Atomism	0	0	0	0
Phenom. thermo.	0	0	0	1

[22] The latter requires extensive historical case studies of experiments and their interpretations regarding predicting novel facts and corroborations.

CHAPTER 4

APPRAISALS OF THE PURE DYNAMICAL VIEW AND COMPUTATIONALISM

This chapter aims to delineate the heuristics of the pure dynamical and the (representationalist) computational views of the mind and argue that the former heuristic is both ad hoc$_3$ and ad hoc$_4$ whereas the latter is neither. To do so, I identify their respective metaphysical hard cores (their negative heuristics), their positive heuristics, whether they are continuous and fact-dependent.[23]

4.1 The appraisal of the pure dynamical view

The *pure dynamicist view* is well represented by van Gelder's (1998) *dynamical hypothesis* (p. 615). According to van Gelder, this hypothesis has two parts. Its first part, the *nature hypothesis*, serves as the hard core of the pure dynamicist research programme; van Gelder (1999) states the following:

> For every kind of cognitive performance exhibited by a natural cognitive agent, there is some quantitative [dynamical] system instantiated by the agent at the highest relevant level of causal organization, such that performances of that kind are behaviors of that system. (p. 9)

The positive heuristic, the second part of van Gelder's dynamical hypothesis, is the *knowledge hypothesis,* the basic claim that one must use dynamical systems theory to model cognition (van Gelder, 1998, p. 9). This claim deserves some unpacking. Although he does not use Lakatos's terminology of research programmes, Chemero (2009) provides the unpacking we need here:[24]

[23] Whether these research programmes are empirically progressive, in the sense defined above, is beyond the scope of this thesis.
[24] van Leeuwen gives Giunti's (1991) "cookbook" for producing such models which is more detailed and technical than the Chemero's recipe, but the main steps are almost identical (van Leeuwen, pp. 299-300).

I. Observe some cognitive activity.

II. Seek "to find the relevant parameters and variables that define the dynamical system that the activity instantiates" (p. 82).

III. Find the "equations that specify the trajectories through the state space defined by the dynamical system" (p. 82).

This heuristic is ad hoc_3, for rather obvious reasons. For any cognitive activity, one must first observe it and then devise equations to model it, where the only constraint is to use differential equations, by deciding upon mapping their variables and parameters to the observations. There is nothing in either the hard core or the positive heuristic which can integrate the various equations devised for the various observed cognitive phenomena. Chemero rightly notes that this is ad hoc because it merely adds yet "another dynamical explanation for another observed phenomenon" (p. 82); e.g., one model for finger wagging, another model for A-not-B task, *etc.*, with different kinds of equations as their central constituents. Indeed, the situation is even worse, because for the same exact data, one can devise various differential equations that will describe it. The heuristic as stated above cannot avoid such under-determination by its hard core or by its positive heuristic. The inevitable result is a "patched up, arbitrary series of disconnected theories" of cognition; hence this heuristic is ad hoc_3. Note further that by the first step of the heuristic, there is a very high level of fact dependency. Moreover, even after successfully "guessing" the form of the equations (i.e., what functions to use in what combination), massive parameter adjusting is required, so that almost every step of the heuristic blocks autonomy. Thus, in addition to being ad hoc_3, this heuristic is also ad hoc_4.

Recall that according to MSRP, creating a model just for a particular phenomenon is not as such defective, because the unit of appraisal is a series of such models. The first two regards in which a model may be ad hoc (whether it predicts other novel facts, and whether any of these are corroborated), are more "local" judgments, in contrast to the third and fourth considerations, which evaluate the cohesion and fact-dependency of the *series* of theories. Accordingly, these latter two considerations are assessed, not when any anomaly arises, but only when its further, sustained accomplishments (or lack thereof) are displayed. When the global accomplishments of the pure dynamical approach are considered in view of these latter two kinds of ad hoc adjusting, the pure dynamical approach is unsatisfactory. These models afford no more than intuition-pumping proofs-of-concepts (as alternatives to representationalism), even when they are empirically progressive. For example, van Gelder's (1999) now classic comparison with the Watt governor, and Thelen and Smith's (1994) explanation of A-not-B error, are of this sort: They unquestionably serve as "eye openers," but the attempts to develop the pure dynamical view into an autonomous research programme – one which avoids being ad hoc in the third or fourth senses – seem quite forlorn from the perspective of MSRP.

In conclusion, I think the pure dynamical view is not a research programme (ad hoc_3) and not autonomous (ad hoc_4). As a series of local models, I find its contributions extremely valuable but as a methodology I agree with Chemero and many others that it is far from satisfactory. Note that because phenomenological thermodynamics is only ad hoc_4 (§3.3), it is methodologically stronger than the pure dynamical view (more on this in §5.1). In the next section I will argue that methodologically, computationalism is stronger than both.[25]

[25] Note that methodological superiority does not entail that one must be committed to it because these are dynamic appraisals. That is, although computationalism is methodologically stronger one can still do research in a purely dynamicist framework. Although here I am defending the methodological superiority of computationalism, I think interactivism is the best research programme for cognitive

4.2 The appraisal of computationalism

There are many different conceptions of computation and hence many different types of computational views of the mind (Rescorla, 2020). For the purposes of this thesis focusing on the *classical representationalist computational theory of mind* (hereafter: “CRCTM”) proposed by Fodor (1975, 1981, 1990) suffices. There are two reasons for this. First, the aim here is to make a methodological contrast with the pure dynamical view using MSRP and not whether all types of computational approaches can be seen as generated by a single heuristic. Second, Chemero (2009) also uses the CRCM for his appraisals and comparisons (pp. 20-22). Hence, I will begin this section by reviewing CRCM, and then give an appraisal of it by making its heuristic explicit.

CRCM has two constituents: *classical computational theory of mind* (hereafter: “CCTM”) and *representational theory of mind* (hereafter: “RTM”) (Rescorla, 2020, p. 13). The intuition behind CCTM is that “mental activity is ‘Turing-style computation’” (Rescorla, 2020, p. 10) where departures from the original Turing machines, including random memory access, parallel computation, and stochastic computations, are allowed. Turing’s (1936) well-known model of computation consists of an ideal machine with unlimited memory and time where the computations are done serially on symbols which are located on an unlimited tape. These computations are done by a central processor and are restricted to the following four: write or erase a symbol in a memory location and move to the next or the previous location. The central processor can enter finitely many *machine states* and which of the four operations it will do depend only on the symbol at the location

science as of now (more on this in §6) .

of the processor and its machine state. Whether this formalism can capture "all humanly executable mechanical procedures over symbolic configurations" (Rescorla, 2020, p. 4) is an open question.

Fodor (1975, 1981) offers a version of CCTM by focusing on the nature of symbols in Turing machines when it comes to minds and offers his RTM (Rescorla, 2020, p. 13). According to RTM, there is a *language of thought* which is constituted by combinations of representations (mental symbols) that have both semantic and syntactic properties.[26] Language of thought is *compositional* in that the meaning of a "sentence" in language of thought depends on the meanings of its parts and the way they are combined. Language of thought is *productive* in the sense that a finite set of primitive representations can be combined in infinitely many ways. Propositional attitude states are relations between the person having such states and language of thought where their semantic properties are based on the semantic properties of the language of thought. There is a *systematicity* of these mental states in the sense that certain propositional attitudes are systematically related to other ones in virtue of, say, their constituent representations. Thus, "mental activity involves Turing-style computation over the language of thought" (Rescorla, 2020, p. 13) and we have CRCTM = CCTM + RTM.

Now, the hard core of this programme may be stated thus: *cognition consists of an intricate complex of representations and Turing-style computational processes on them*. Its positive heuristic can be modeled on Clark's (1976) statement of the positive heuristic of the atomist programme in thermodynamics (p. 45):

I. Make simplifying assumptions about the task at hand, together with the nature of representations and their processing, to devise algorithms

[26] Their casual role, though, depends only on their syntactical properties.

operating on representations able to perform the target task by executing the specified computations.

II. These simplifying assumptions should be weakened or even replaced to better approximate the real phenomenon under consideration, *e.g.*, a "real" A-not-B task, a cleaning agent in a "real" office, a decision task in a "real" imprecise context, *etc*.

III. Using the specific theory thus created, its consequences should be investigated to obtain results regarding the internal properties of the phenomenon under investigation and to derive the known observable characteristics, such as surface level language, folk psychology, or self-consciousness.

This heuristic is neither ad hoc_3 nor ad hoc_4 and the reasons for this are like those regarding the atomists' heuristic. Rather than well-established Newtonian laws of mechanics, computational representationalism invokes a well-established theory of computation.[27] For this reason alone the hard core of computationalism is sufficiently constrained to provide a unifying framework. Thus, any theory generated by using this heuristic can, in principle, be traced back to a common abstract computational paradigm. Hence the series of computationalist theories are well-connected, hence not ad hoc_3; computationalism suffices as a research *programme*.

Regarding autonomy, before considering any empirical content, computationalism affords predictions depending upon the assumptions made in Step 1. For example, the more complicated the processes, the more computational time is required to accomplish a given task. Hence one can explain certain experimental

[27] Indeed, it is tempting to push this analogy further: Obtain the computationalist heuristic by replacing "representations" with "atoms" and "Turing's theory" with "Newton's mechanics" in the atomist heuristic (together with other minor modifications).

results, or can design an experiment to obtain relevant results, to test this assumption, *e.g.,* mental rotations is one such case (Chemero, 2009, p. 81). Similarly, changing assumptions and determining their consequences can also proceed largely by internal (hence autonomous) theoretical requirements. For example, change the assumption that processes on representations are probabilistic rather than deterministic, and try to determine what this may imply for cognition. This has of course been done; we have Bayesian models and probabilistic robotics (Thrun et al., 2006). One can keep the processes deterministic and instead change the content of representations from *objectivist* to *action-oriented representations* to work theoretically on what such a change entails for the computational architecture (Clark, 2001, pp. 131-133). Since all of these are possible with relative autonomy from empirical data and from anomalies, this heuristic is not fact dependent and thus is not ad hoc[4].

The methodological weakness of the pure dynamical approach and the strength of computational views, though not cast in MSRP terminology, are not new debates within philosophy of cognitive science. My aim in this chapter was to show that MSRP suffices to account for and to augment the terms of these debates. As we will see, Chemero reached similar conclusions regarding their methodological powers. To remedy this situation, he supplements the pure dynamical approach with additional requirements on its modelling norms. The next chapter summarizes his approach and gives separate appraisals for his two proposals.

CHAPTER 5

APPRAISAL OF CHEMERO'S RESEARCH PROGRAMME

We have seen in the previous chapter that MSRP conforms to the received wisdom regarding the methodological evaluations of pure dynamical modelling and computationalism. In this chapter I will argue that it can differentiate more subtle methodological differences by giving an appraisal of Chemero's (2009) radical embodied cognitive science. The first section delineates some issues regarding Chemero's understanding of the debate between the physicists in the 19th century and Newtonian mechanics itself. The next two sections will be the appraisal of his methodological proposals proper. The chapter concludes with some observations regarding realism.

5.1 A reductive interpretation of the history of physics? And of physics?

In this section I will argue that there are reasons to suspect that Chemero has a reductive viewpoint when it comes to the debate of 19th century physics and Newton's programme. The importance of the latter stems from the fact that the programme of atomistic physics is basically a variety of Newton's programme. Thus, a reductive understanding in Newton's programme might trigger a chain of reductions along the way. This in turn is important for this thesis because such a reductive reading either as a cause or a symptom has direct relevance to his claims that his methodology is as strong as the atomists of the 19th century. Since he does not discuss much about Newton, my claims are mainly worries constructed from scattered remarks of his. When it comes to the 19th century debate, his position is clearer and hence one can make correspondingly stronger claims.

Recall that Chemero (2009), unlike Clark (1976), mostly focuses on the debate between Mach and Boltzmann which corresponds roughly to the last stage of the debate. This last stage includes discussions regarding realism versus instrumentalism, which suits Chemero's agenda. But we have seen that the debate is much broader (§3.3) and the issue of realism became a serious contender only later at the degenerating phase of the kinetic programme (and that with an intuitive understanding, by some of the leading scientists, that the heuristic of the phenomenological thermodynamics was weak - see §5.4). Not taking into consideration the early periods of the debate as well as the later progression after Mach and Boltzmann might have enabled Chemero (2009) to claim that Boltzmann's guide to discovery argument "was devastating to the phenomenalist picture of physics" (p. 79, see also Chemero, 2000, p. 640). But this is clearly wrong according to the methodology of research programmes. We have seen that, on the contrary, it was rational for those scientists to choose whatever side they saw fit. If anything was "devastating" to the adherents of phenomenological thermodynamics, it was not Boltzmann's guide to discovery argument but Einstein's 1905 paper explaining Brownian motion within the atomistic heuristic (Clark, 1976, pp.93-98).

Chemero’s overestimation of the "guide to discovery" argument is accompanied by his underestimation of phenomenological physics based upon an analogy between its methodology and the dynamical stance. Indeed, Chemero (2009) says that the disconnected models of the dynamical stance are "just as in the case of phenomenological physics" (p. 82). And it gets worse, because of the following:

> Boltzmann’s guide to discovery argument criticized phenomenological physics because it had no means of predicting as-yet-unobserved phenomena to test and extend it as a theory. We can see now that the real problem that Boltzmann pointed to is that phenomenological physics is not a theory at all, it is a set of models created on the fly to be applied to incoming data. (Chemero, 2009, p. 100)

But a close reading of Clark (1976) and Boltzmann (1900) suggests that both claims are wrong. As for the claim that phenomenological physics had no novel consequences, Clark (1976) gives detailed explanations of a couple of them and states explicitly that from "each particular theory (Carnot's theory, the mechanical theory, and phenomenological thermodynamics) a sequence of novel predictions followed which were subsequently confirmed" (p. 74). Also, Boltzmann (1900) states that with the help of Clausius' entropy and some further work of Gibbs regarding thermodynamic potentials "the most surprising results were reached in the most varied fields, as in chemistry, capillarity, etc." (p. 246). Thus, the issue is not whether there are novel predictions, but in what manner they are obtained.

As for the second claim stating that phenomenological thermodynamics is basically a set of models, which is also his critique of the dynamical stance, the results of §3.3 and §4 show that, on the contrary, the heuristic of phenomenological thermodynamics is stronger than the heuristic of the dynamical stance; it is a research programme. Recall that fact dependency was a result of the contentless nature of the two universal laws (conservation of energy and entropy increase) as formulated by Planck. The contents were obtained from the empirical situation at hand. The crucial point here is that the heuristic provided a principled way of providing exactly when during the derivations the input from the constitutive equations were required. Such constraints are simply missing in the heuristic of the dynamical stance. So, emphasizing a similarity of an aspect of phenomenological physics to dynamical stance, namely both being fact dependent, harbors the risk of reduction of the whole of the former's methodology to the latter's.

As for Chemero's reductive interpretation of the Newtonian programme, note that while explaining van Gelder's account of the Watt governor Chemero (2009) says "*just as Newton did* [emphasis added] in his descriptions of the physical world, the behavior of the system of interest is observed, and mathematical equations that describe that behavior are found" (p. 69). He then gives the equation of the governor explicitly which is as follows[28] (van Gelder, 1999, p. 5):

$$\frac{d^2\theta}{dt^2} = n^2\omega^2 \cos\theta \sin\theta - \frac{g}{l}\sin\theta - r\frac{d\theta}{dt}$$

Now the warning sign for me is that this is not "just as" in Newtonian dynamics, this *is* Newtonian dynamics itself and the equation *is* Newton's equation itself for Watt's mechanical apparatus.[29] It is an excellent example of a *derivation* of an equation using Newton's heuristic (Beltrami, 1987, pp. 152-55) and is far from any data fitting or "attractor hunting" (see the next section for more on these extensions of HKB+).

After finishing the dynamical explanation of the Watt governor, Chemero (2009) says the following:

> In this, as in all dynamical explanations, once we have found equations such as these for the Watt governor, it is agreed that we have explained the Watt governor's behavior; we have a perfectly general, counter-factual supporting description of its behavior, as is provided in Newtonian physics. (p. 70)

This is true but only because of a strong heuristic backing it up so that we have the general principles for deriving this and many other equations in an autonomous manner. Because there is no discussion of how that equation is obtained or its Newtonian characteristics (e.g., the existence of the second derivative or gravitational and centrifugal forces), Chemero seems to be focusing solely on the result and the properties there of, such as being general or counter-factual supporting

[28] Note that in both Chemero (2000) and Chemero (2009) there is a typo in the equation, namely, instead of n^2 we just have n.

[29] For a trained eye this is almost immediate: Newton's fundamental equation $F = ma$ takes the form of $F_1 + F_2 + F_3 = ma$ where the mass term m cancels in the above equation because both sides depend equally on it. The acceleration a becomes the angular acceleration and hence there is a second derivative of θ (note that the order is reversed for the equation in the text, viz., $ma = F_1 + F_2 + F_3$ where the forces are centrifugal, gravitational, and friction forces respectively).

description of the behavior. For me this is a big worry because therein lies the real strength of the equation, and not in the fact that it is "general" or makes "counter-factual supporting" descriptions. Such focus on the result of a strictly Newtonian derivation will in turn let one to bestow to any data fitting equation a surface level similarity with it and hence a methodological strength that seems to be on a par with a strong heuristic (like HKB's extensions and applications seems to be, as we will see in the next section).

Admittedly these might not be much of knock down arguments for showing Chemero's reductive approach to the methodology of Newtonian physics. It is more of an attempt to make explicit his value attributions in his philosophy of science. Following Clark's (1976) remark that to understand working scientists, what is important is not what they say, "but what principles they *adopted and followed in practice* in developing their theories" (p. 44), more evidence can be harnessed for these claims by investigating Chemero's defense of HKB and the way it provides having a guide to discovery. This will reveal that it cannot be on a par with the methodology of atomism or computational cognitive science.

5.2 Appraisal of the HKB+ case

We have seen in §2.2 that Chemero establishes norms that any candidate methodology must satisfy, at least to some degree. These norms are providing a guide to discovery, being as extensive as possible, and being a realist theory (Chemero, 2009, p. 97). He claims that if a research programme provides a guide to discovery then it is methodologically on a par with computationalism or atomism (Chemero, 2009, p. 85). Therefore, since his model-based proposal, namely HKB+, guides discovery, it is on a par with computationalism. Below I argue that his model-based methodology satisfies at most MSRP's requirement of continuous growth, and his account of guidance to discovery renders this norm so weak that it can be attributed to both the model-based and to the computationalist methodology. On the other hand, we have seen that computationalism satisfies not only MSRP's requirement of continuous growth, but also that of *autonomous* continuous growth. Therefore, Chemero's norms lack sufficient discrimination to differentiate between these two types of methodologies. The key issue is not whether a guide to discovery exists but *the way* it points to new hypotheses and experiments.

The appraisal of HKB+ can be given by first identifying its hard core together with its positive heuristic, and then assessing whether and what kinds of ad hoc-ness are present. Instead of this direct argument, I will give an indirect one arguing that no matter what the heuristic is, if the resulting research type of HKB+ is as described by Chemero, it can at most be ad hoc_4. This will also demonstrate the flexibility of MSRP when it comes to types of arguments one can use when assessing particular fields.

According to Chemero's earlier work, HKB by itself is “not sufficient to save dynamical cognitive science from fact dependence-it still provides no “guide to discovery”” (Chemero, 2000, p. 645).[30] It, however, begins to provide such guidance

[30] One important issue regarding HKB itself for the purposes of this thesis is that the HKB’s original paper spends a considerable time deriving the potential equation from equations of the individual hands and their coupling (Haken et al., 1985, pp. 351-353). These equations are nonlinear, dissipative oscillator equations and such a derivation implies a connection with the rest of the physics, all the way to far from equilibrium thermodynamics. This in turn means that a hope for continuity exists, at least in principle. But as we have seen in the penultimate paragraph of §2.3.1 Chemero (2009) gives only a passing remark about this aspect of HKB and later, after giving the extensions of HKB that are based on the modifications of the potential function, says that “the vague suggestion concerning what coordinated structures *are* bore fruit in more concrete suggestions concerning how such structures were to be *modeled*” (p. 96). I take it that this shows how different methodological viewpoints result in focusing on different stages of a particular research. For me, the possibility of continuity and autonomy offered by *deriving* the potential function from more fundamental physical considerations is much more satisfactory than the one obtained by simply beginning with a potential function (which can be obtained by pure data fitting) and investigating its possible modifications and extensions. Note also that this latter aspect is very similar to what Boltzmann (1900) calls the “mathematical phenomenology” whose methodology is even weaker than the “general phenomenology” (p. 250).

thanks to the work of Kaipainen and Port (unpublished) who have developed a "generalization" of the HKB to speech actions. They did this by adding to the two oscillatory functions of HKB another oscillatory function to account for the additional attractors in their speech action data.[31] The behaviour of the equation then fits the data. Therefore, Chemero (2000) anticipates, it "will allow *non*-ad hoc extensions to our understanding of cognition. It does so by potentially capturing all of the rhythmic behavior with one dynamical model" (p. 645). What differentiates this from other (mere) data-fitting modeling is that it is a "rich and extensible unifying model" (Chemero, 2009, p. 95). On the other hand, Chemero's (2009) two characteristics of being fact dependent are making empirical generalizations and altering the equations to fit anomalies in the experimental results (p. 80). Note that Kaipainen and Port's (unpublished) extension is done in just this manner. They made an empirical generalization from the oscillatory motion of fingers to periodic patterns for speech production, and then altered the original equations to fit the data. This is a fact dependent method and the result being rich and extensible cannot change this.

Some other extensions of the basic HKB model are made similarly. For example, the extension to cover asymmetries resulting from handedness is also achieved by adding two more oscillatory functions (Chemero, 2009, pp. 91-92). The mathematical nature of the modifications Chemero (2009) defends are like Ptolemy's programme. Recall its ad hoc$_4$ heuristic from §3.3: save the phenomena by a combination of uniform periodic motions. Thus, future historians of cognitive science may interpret the way the new models are generated in HKB+ as *the Ptolemean phase* of cognitive science.

[31] That is to $V = -acos\varphi - bcos2\varphi$, they added another faster cosine so that the new $V' = -acos\varphi - bcos2\varphi - ccos3\varphi$.

Indeed, the fact dependency of these cases must hold for any extension of HKB, since these extensions are made in ways *not* prescribed by the model itself, nor from Chemero's norms but is, so to speak, user-designed. That is, it is *necessary in principle* that we must first observe and decide what is oscillatory and what is not, and then apply the model's equations with suitable modifications. This is a major obstacle to autonomy and therefore HKB+ is ad hoc_4. An autonomous heuristic reverses the order, namely, instead of deciding first whether two phenomena are similar and then modifying a set of equations from one to fit another, the heuristic itself shows us, and *surprises us* that such diverse phenomena results from a common mechanism. Therefore "being as extensive as possible" cannot be a constituting norm of a mature, unifying methodology; instead, such extensibility must be a consequence of it.

There is also an ambiguity regarding the distinction between extensions and applications; this issue concerns ad hoc_3-ness, that is, merely patched-up theorizing. In Chemero's earlier work there is an ambiguity regarding the distinction between guide to discovery and being as extensive as possible (Chemero, 2000). It seems that his initial claim about being as extensive as possible implies guiding to discovery for he says that dynamical cognitive science could "provide a guide to discovery, if it were able to posit a generally applicable type of dynamical model that accounts for a wide range of cognitive phenomena" (Chemero, 2000, p. 643). I think it is more appropriate to call some of the "extensions" of HKB merely "applications with modifications." For example, according to Chemero (2009), there is an "extension" to learning where "learning is understood as a phase transition" (p. 89). Similarly, we have seen that there is another "extension" to representation-hungry gear problems. Since the immense expressive powers of phase transitions implies that almost anything can be modeled with them,[32] there is a risk of confounding mere applications with genuine extensions.

To see this, note that changing the parameters of a differential equation leads to qualitative changes in its *attractor landscape*. Moreover, the possible types and ranges of these changes are unbounded (See Abraham and Shaw (1987) for stunning visual depictions of some phase transitions and self-organizations). This means that the modeled phenomena will also show qualitatively different behaviors under the same set of rules (i.e., the form of the differential equations remain the same, only the parameters' values change). For example, given a dynamical model of an aspect of speech production, it is almost certain that some other phenomena, say, an aspect of economics, will show a qualitatively similar behaviour. Then, the quantitative mismatches can be taken care of by adding small corrective terms to the equations. Now, is this a legitimate "extension" of the speech production to economics? If so, what reason in principle blocks the converse, namely, an "extension" from economics to speech production?

An actual example of this from cognitive science can be given: Aerts (2007) applied the mathematics of quantum mechanics to conjunction of concepts and found that his model fit the data. The experiments showed the interesting effect that subjects might assign higher values for memberships to conjunctions than its constituents. For example, Guppy has a higher value of membership to the *Pet-Fish* category than either to the Pet category or to the Fish category (p. 2). Given such a fit, are we justified in saying that quantum mechanics found a genuine extension to concept research? This is unlikely. Rather, what is likely is that an *aspect* of the mathematical formalism of quantum mechanics finds an *application* in a different

---

[32] Including modeling Turing machines (Asarin & Maler, 1994). Note also that dynamical systems have more power than Turing machines (Bournez & Cosnard, 1996).

area, namely, an area where the influence of context has non-negligible effects (I believe Aerts claims so much). Any claim stronger than this will be ad hoc_3.

Therefore, following the norm of being as extensive as possible in such an unconstrained manner will at best produce a disconnected series of theories, namely, an ad hoc_3 methodology. Nevertheless, this issue is not fatal to HKB+ because by restricting the scope of its extensions, for example only to musculoskeletal phenomena (instead of, say, to representation hungry problems), it can be saved from being ad hoc_3. Granting this, its only defect will be its ad hoc_4 heuristic. Thus, Chemero's (2009) claim that his model-based proposal is on a methodological par with computationalism is false because the heuristic of computationalism is not ad hoc_4 (§4). A summary of the appraisals of different methodologies discussed so far are given in Table 2.

Table 2. Appraisals Given So Far

Appraisal at a non-degenerating phase of a given field	ad hoc_1 (theory does not also explain other novel facts)	ad hoc_2 (none of the novel facts are corroborated)	ad hoc_3 (disconnected series)	ad hoc_4 (not autonomous)
Atomism	0	0	0	0
Phenom. thermo.	0	0	0	1
Pure Dynamical	0	0	1	1
Computationalism	0	0	0	0
HKB+ (under a charitable reading)	0	0	0	1

5.3 Appraisal of the Gibsonian case

In this section I will give an appraisal of Chemero's (2009) radical embodied cognitive science proper by making explicit its hard core and its positive heuristic. Its hard core can be stated as follows:

Cognition consists of anti-representational processes whose primary constituents are affordances which obey the principles of Gibsonian ecological psychology.

Its positive heuristic can be formulated as follows:

I. Make assumptions regarding what affordances are in each situation.

II. Make simplifying assumptions regarding affordances in the given situation so that a set of differential equations can be written down. These equations will model the coupling between the organism and the environment.

III. The above two simplifying assumptions should be weakened or even eliminated to better approach the real phenomena under consideration.

IV. Depending on the match between the model and the experimental data, either make changes to the model (e.g., add terms to the equations) or change the assumptions made in 1 and 2.

Note that I have tried to render the hard core of Chemero’s heuristic as close as possible to the heuristic of the atomist program, where the affordances in the former play a role similar to the atoms in the latter. The reason for this is to have an appraisal as charitable as possible. Another motivation for this was the following. As we will see below, the modifications of affordances from affordances 1.0 to 2.0 is similar to what atomists were doing with their modifications regarding the nature of atoms as prescribed by their heuristic (§3.3). As we will see, the similarities stop here.

Since affordances are central for this heuristic and since their detailed account was not reviewed in §2.3.2, I will present a more detailed review of the series of definitions beginning with what Chemero (2009) calls "affordances 1.0" (p. 136). These are the class of definitions that take affordances to be "properties of the environment that have some significance to some animal's behaviour" (p. 137). There are two possible readings of this and correspondingly two camps within the affordances 1.0 class. The first camp, as exemplified by Reed (1996), takes affordances to be organism independent properties of the environment that can create selection pressure (Chemero, 2009, p. 137). Organisms evolve to perceive and use them (Chemero, 2009, p. 137). The second camp, as exemplified by Turvey (1992) and many others, see affordances as *dispositional properties* of the environment (Chemero, 2009, p. 137). To have a dispositional property is to have "tendencies to manifest some other property in certain circumstances" (Chemero, 2009, p. 137). The example Chemero (2009) gives is being fragile. Glass is fragile but this manifests itself only around sharp or hard entities (p. 137). Such actualizing circumstances had to be paired with a property of the organism for them to be affordances. Whether the organism's abilities or body scale or something else is to be counted as part of an actualizing circumstance is a matter of debate. Chemero is against the idea that affordances are some sort of property in the environment (p. 139). Instead, they must be features of the whole environment and this environment includes the animals too. Features are context dependent in that they refer to a *situation* here and now which might never happen again (p. 140). Hence, a new definition that incorporates such insights is necessary.

Affordances 1.1 are defined as *relations* between *abilities* and *features* (Chemero, 2009, p. 145). Here, abilities are different from dispositions that were used in the definition of affordances 1.0 because dispositions lack normativity. Abilities are functions that can malfunction. *Features* are not properties of the objects in the environment or of the organism itself but something that belongs to both the organism and the environment. It lets the organism perceive "that the situation as a whole has a certain feature, that the situation as a whole supports (perhaps demands) a certain kind of action" (p. 140). For example, there are experiments showing that body scale is not important when presented with stair-climbing affordances (p. 143). By comparing children, young adults, and older adults, it was found that subjects in such experiments "perceive stair-climbing and descending affordances not as the ratio between leg length and riser height … but rather as a relation between stepping ability and riser height"[33] (p. 143).

Moreover, as a relation between abilities and features, affordances 1.1 are static, yet what we need is an understanding which is dynamic (Chemero, 2009, p. 150). Thus, affordances 2.0 is offered to consider the real time dynamics of the organism with its environment. They are obtained by modifying affordances 1.1's relational character by "considering the interaction over time between an animal's sensorimotor abilities, that is, its embodied capacities for perception and action, and its niche, that is, the set of affordances available to it" (Chemero, 2009, p. 150). Unlike affordances 1.1, he does not give a clear definition of affordances 2.0 but the following features clarifies what he is aiming at.

First, this is "a variety of niche construction" but over shorter time scales as compared to whole populations or species (Chemero, 2009, p. 151). Second, it makes it possible to make a fusion with the *enactivists* explicit (Chemero, 2009, p. 152).

---

[33] A possible issue here is that riser height is a property of something in the environment and *not* a feature. Although Chemero discusses and justifies the ability relata of the affordances in these experiments, he does not address why riser height is a feature. In general, an account of properties in a feature based ontology seems to me non-trivial and hence must be addressed.

These two points do not affect the arguments in this thesis. Third, and this is the relevant one, Chemero (2009) makes the following remark about this shift to affordances 2.0:

> This is not so much a new way of understanding affordances as a critique of prior attempts to come up with a *definition* of the term "affordance." Ecological psychologists have always been aware of, indeed keenly interested in, the interaction of affordances and abilities in real time…. Affordances 2.0 is an attempt to develop a theoretical understanding of affordances that is more in line with the experimental and explanatory practices of ecological psychologists. (p. 151)

The only difference between affordance 1.1 seems to be the incorporation of a temporal dimension, albeit in a *post hoc* manner because the nature of affordances are as before. We just let affordances and abilities to "causally interact in real time" and be "causally dependent on one another" (Chemero, 2009, p. 151). At each instant we have the same definition of what an affordance is (an instance of an affordance 1.1) but with affordances 2.0, we have a local and fast niche construction.[34] So, this new definition does not change the nature of affordances but considers its development over an intermediate time scale. This in turn means that it will still not be sufficient to rescue Chemero's methodology as we will see below.

Given these definitions of affordances, Chemero (2009) claims that his Gibsonian theory-based methodology is not fact dependent because it can provide to dynamical systems approach what atomism provided to thermodynamics (p. 83). Therefore, his research programme must have powers equivalent to a research programme which is neither ad hoc$_3$ nor ad hoc$_4$. My claim is that it is both. The structure of the argument will be as follows. By ramifying Clark's notion of fact dependency, I will delineate additional necessary conditions for any autonomous research program. I will then argue that Chemero's utilization of Gibsonian

[34] Local and fast as compared to a population of organisms and their niche constructions (Chemero, 2009, p. 152).

methodology for the radical embodied cognitive science[35] has these dependencies so that it is both ad hoc_3 and ad hoc_4. Perhaps surprisingly this implies that it is worse than HKB+ (as HKB+ was *not* ad hoc_3 under a charitable reading (§5.2)).

Recall from the previous section (§5.2) that one reason for HKB+'s methodology to be ad hoc_4 was deciding *first* what surface level phenomena is oscillatory and *then* proceeding with a model. Similarly, in the Gibsonian case, the first step in its heuristic demands one to decide what affordances are in each situation. Therefore, such an ontological step in the heuristic of the Gibsonian case will also render autonomy impossible. Thus, its methodology is ad hoc_4. Call this "phenomenal fact dependency."

But the situation in the Gibsonian case is worse than the HKB+ case because in the latter the decision's base, namely being oscillatory, is restrictive and precise. That is in deciding what surface phenomena to be oscillatory there are two constituents: first the base notion, namely, the concept of oscillation, and second the phenomena to be explained using the base notion, namely the predication of the property of having oscillatory constituents. However, in the Gibsonian case this first relatum, namely the definition of an affordance, is not as restrictive. Recall that an affordance is a dynamic relation between an *ability* of the organism and a *feature* of the environment. Even though this in itself might not be a major defect of Chemero's programme,[36] it nevertheless poses serious issues when it comes to heuristic power. The reason for this is that all such definitions within a strict Gibsonian framework will be too broad because the environment is too broad. Furthermore, even if this wide scope was not problematic, there is no in principle cohesion between these

[35] Hereafter "Gibsonian" will be used in the sense of Chemero's interpretation and utilization of Gibson's framework, *not* in the sense of Gibson's own views.
[36] Note that progress is possible even with inconsistent foundations (for example, Bohr's attempts in the beginnings of quantum mechanics (Lakatos , 1970, p. 140)).

affordances. The examples of affordances in Chemero (2009) include such diverse kinds as the riser height perception (p. 143), inertial tensor perception (p. 154), and direct entropy perception (p.127). Depending on the situation and the researcher, various qualitatively different theories within the same research programme will emerge depending on the assumptions of what affordances are. Call this kind of fact dependency "ontological fact dependency." To sum, both phenomenal and ontological fact dependencies arise when there is a necessary fact dependent decision to make to initiate your heuristic.

One might counter this by noting that even atomists tried switching from a classical atom to a force field model of Maxwell. And hence radical embodied cognitive scientists should be allowed to switch from an oscillatory base to an affordance base. But Maxwell is explicit that one reason for his vortex model was to prevent arbitrariness:

> The success in explaining phomena [sic] does not depend on the ingenuity with which its contrivers 'save appearances' by introducing first one hypothetical force and then another. When the vortex atom is first set in motion all its properties are absolutely fixed and determined by the laws of motion of the primitive fluid, which are fully expressed in the fundamental equations. (Maxwell, 1876, pp. 471-472, as cited in Clark, 1976, p. 85)

Following Maxwell, shouldn't we also say that the success in explaining cognition should not depend on the ingenuity with which its practitioners "save appearances" by introducing first one affordance and then another depending on the situation?[37] This in turn will lead to different kinds of differential equations depending on the

[37] Note that although Maxwell's vortex atom model was theoretically within the atomistic heuristic, its mathematical complexities were impeding progress. Thus, it was eventually abandoned (Clark, 1976, pp. 84-85, see also footnote 195). Also, Boltzmann (1900) suggests that such a technical difficulty entails a distance from reality: "their structure took so complicated a form that they could not possibly lay claim to any other validity than … that of ideal mechanisms producing effects in some way analogous to the phenomena of reality, certainly not that of definitive facsimile representations of what actually took place in nature" (p. 249).

affordances and hence to different kinds of models. Thus, radical embodied cognitive science has no inherent constraints which can prevent it from generating disconnected series of models. This means that the Gibsonian heuristic is ad hoc$_3$.

The ontological and the phenomenal fact dependencies are independent in the sense that one might have phenomenal but not ontological fact dependency. For example, the HKB+ case under the charitable reading is like this. Conversely, one might have ontological but not phenomenal fact dependency, that is, there may be no need to decide what in the environment corresponds to your fundamental entities, but those fundamental entities themselves are vague. Perhaps the early beginnings of quantum mechanics were like this, but this point requires historical analysis which is beyond the scope of this thesis.

Now, a crucial difference exists between the hard cores of the atomists and radical embodied cognitive scientists. The hard core of the atomists refers explicitly to Newtonian mechanics, which is already a well-defined and strong research programme. An important side question here is what were the reasons for the possibility of a consensus in the atomist community be it the standard atom model or the vortex model as its core? One reason for this might be that our everyday intuitions are well suited for particles or waves to the exclusion of any other ontology. Another reason might be that even in cases where everyday intuitions fail,[38] strict mathematical formalism enables a consensus to exist. This in turn puts strong constraints on stretching the auxiliary assumptions of the heuristic of such a research programme. Thus, the resulting models have a lesser chance of being disconnected. But, when it comes to cognition, the non-existence of such constraints creates a fundamental instability for any programme that takes affordances

[38] For example, as in quantum mechanics.

themselves as the basic entities of research. A possible solution to this issue might be taking affordances as derived from conceptually more basic entities/processes rather than taking them as the fundamental building blocks of a research programme. Such an approach is clearly absent in radical embodied cognitive science.

A comparison between the heuristics of the Gibsonian case and phenomenological thermodynamics reveals further issues regarding the fact dependencies of the Gibsonian case. At best it can be just as good as phenomenological thermodynamics. To see this, assume that there are no scope issues regarding the definition of affordances and no decision issues regarding what affordances are in each situation. That is, assume there are no ontological and phenomenological fact dependencies. But, because affordances are relational, the particular form of that relation, the content of the relation must be provided by a low-level empirical law. Note that this is now becoming exactly like the situation for phenomenological thermodynamics. Recall that the reason for such a necessity for a low-level empirical law was the contentless nature of the two fundamental laws (§3.3). A typical recourse in such cases is the ideal gas law: $PV = NkT$. Here, P is pressure, V is volume, N is the number of molecules, k is the Boltzmann constant, and T is the temperature. To derive the consequences of the two fundamental laws, energy and entropy need to be calculated. That means the work done *by* the system or *on* the system must be calculated. This usually involves a relation between V and P which is provided by the ideal gas law (Volkenstein, 1986/2009). But note that there are no issues here regarding the definitions of pressure or volume, and there are no issues regarding deciding what pressure or volume is in each situation. Hence the fact dependency of phenomenological thermodynamics is of a different kind than the fact dependencies of the Gibsonian case we have seen so far. Let us call this "epistemological fact dependency." Note that epistemological fact dependency is just Clark's (1976) "fact dependency." Now, the Gibsonian case not only suffers from all 3 of the fact dependencies, but the best it can ever do is to suffer from epistemological fact dependency. That is, the best it can do is to be on a par with phenomenological thermodynamics. Thus, it is an ad hoc$_4$ heuristic and Chemero's (2009) claim that his methodological proposal will provide to dynamical stance what atomism did to phenomenal thermodynamics is wrong[39] (Table 3 below provides a summary of the appraisals given so far).

Table 3. Appraisals Given So Far

Appraisal at a non-degenerating phase of a given field	ad hoc$_1$ (theory does not also explain other novel facts)	ad hoc$_2$ (none of the novel facts are corroborated)	ad hoc$_3$ (disconnected series)	ad hoc$_4$ (not autonomous)
Atomism	0	0	0	0
Phenom. thermo.	0	0	0	1
Pure Dynamical	0	0	1	1
Computationalism	0	0	0	0
HKB+ (under a charitable reading)	0	0	0	1
Gibsonian Case	0	0	1	1

[39] What happens if we force the analogy of the atomist heuristic and the Gibsonian case? Having a single kind of entity (i.e., "affordances") resembles the atomist heuristic (i.e., "atoms"). But the Gibsonian case is like postulating atoms without Newton's mechanics. This in turn would mean that there is no principled reason to block a conclusion stating there are as many kinds of atoms as there are different kinds of phenomena out there because *everything* affords something. The main problem with affordances is that every empirical situation has affordances themselves, and not a manifestation of affordances' different combinations according to fixed rules. Atoms, on the other hand, do exactly this, namely, all the diverse phenomena we observe are manifestations of atoms in different combinations under fixed rules (say, Newton's equation in mechanics or Schrödinger's equation in quantum mechanics)

Figure 1 below shows the relation between fact dependencies and ad hoc-ness. Note that this figure is merely a summary of the distinctions introduced so far and is not exhaustive. For example, there might be fact dependencies different than the ones shown in the figure which can cause ad hoc_3 or ad hoc_4 methodologies. One can see from the figure that computationalism and atomism do not have the depicted fact dependencies. In contrast, if a methodology has one or more of the fact dependencies, then it will have one or more of the ad hoc-nesses labeled by the arrows in the figure.

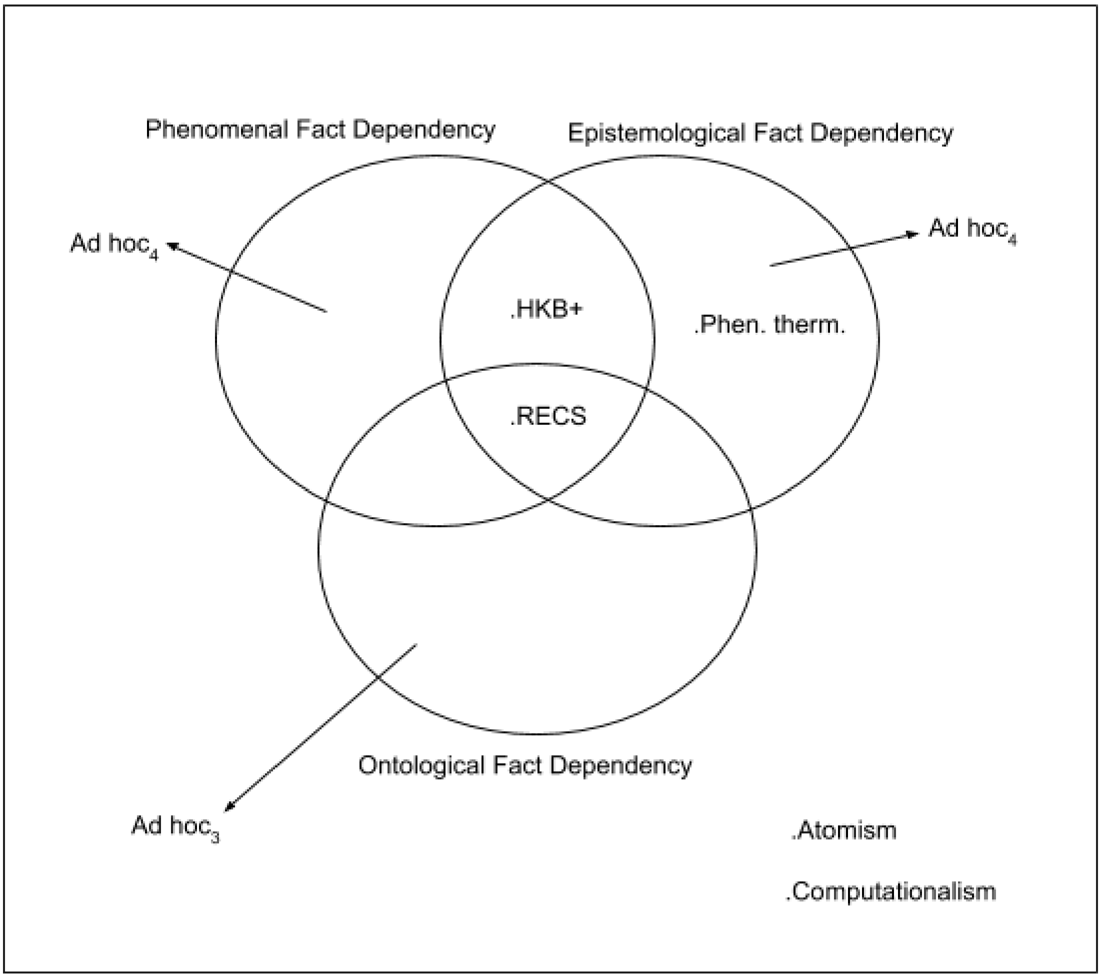

Figure 1. Types of fact dependencies and ad hoc-ness ("RECS" stands for "radical embodied cognitive science")

Regarding radical embodied cognitive science's utilization of dynamical systems, the following summary given by Chemero (2009) regarding his programme is conclusive:

> Here, then, is radical embodied cognitive science: Animals are active perceivers of and actors in an information-rich environment, and some of the information in the environment, the information to which animals are especially attuned, is about affordances. Unified animal–environment systems are to be modeled using the tools of dynamical systems theory. There is no need to posit representations of the environment inside the animal (or computations thereupon) because animals and environments are taken, both in theory and models, to be coupled. (p. 160)

The use of dynamical systems tools to model Gibsonian insights in the same way as they are used in the (pure) dynamical stance brings up another issue. Can one use the same mathematical machinery but come up with a different interpretation so that one ends up rendering a weak heuristic strong? For this to be possible, the interpretation must constrain the methodology in such a way that the external dependence should decrease so that autonomy can increase. Chemero's use of Gibson as a background theory does not change the modeling practices of the pure dynamical stance so it cannot accomplish this feat. To make this point clearer, perhaps one should approach the issue from the opposite direction. Take any purely dynamical model. One can then come up with a Gibsonian explanation almost by naming: Declare the relevant parameters in the model as affordances! The issue is that this is precisely what *cannot* be done in, say, a Newtonian setting. The differential equation written down must have the form $F = ma$ or must be derivable from it (as we have seen with the equation of Watt's governor in §5.1).

In terms of MSRP, this question can be formulated more generally via MSRP's distinction between the negative and the positive heuristic of a research field: Is it possible to keep the positive heuristic fixed while changing the hard core? If yes, can such a move render an ad hoc$_i$ heuristic to an ad hoc$_j$ heuristic, or even completely get rid of ad hoc-ness? In dynamical accounts, it seems that many researchers are trying to defend their adherence to the mathematics of dynamical systems in such a manner. There is no guarantee that keeping the way dynamical systems are used in the positive heuristic while changing the hard core will work as a remedy for methodological issues. For example, declaring that cognition is a "quantitative" dynamical system (van Gelder, 1999) or offering Gibsonian principles as a hard core (Chemero, 2000, 2009) do not suffice for eliminating ad hoc-ness given the fact that the way these methodologies use dynamical systems in their positive heuristic is not that different from pure dynamical modelling.

One final remark is the following. Recall from §3.3 that atomists had a strong heuristic power whereas thermodynamicists had more empirical power. There is also a similarity here with radical embodied cognitive science. Writing down differential equations for the particular cognitive task, and then modifying them will almost instantly create empirical success. In contrast, the distance to empirical phenomena is bigger especially if one considers entities that are internal (e.g., representations) so that their surface level manifestations and interactions require derivations one way or the other. This thesis tries to argue that empirical power should not be reified into having a strong heuristic.

5.4 Some observations on realism

I have mentioned that realism is an important issue for Chemero (§2.3). It is not so for MSRP. But, during the later phases of the debate between the atomists and the phenomenologists, realism did become an issue for these programmes. This happened either at a degenerating phase of the research programme (e.g., atomism) or after the realization that a weak heuristic is blocking further progress (e.g., phenomenological thermodynamics). At such a juncture the physicists of the two camps tried to switch either from a realist position to an instrumentalist position (e.g., atomists) or from a formal mathematical position to a realist position (e.g., phenomenological thermodynamicists).

For example, Boltzmann, being aware of the degenerating phase of the atomistic programme, changed his philosophy from a strongly realistic stance of atoms to "a methodology based upon regarding theories as mental pictures or classificatory devices to be compared as to their fruitfulness and to their ability to provide economical classifications" (Clark, 1976, pp. 89-90). In the rival camp and in the opposite direction, Ostwald, being aware of the weak heuristic of thermodynamics, tried to give "an ontological basis" to the programme by taking energy as the fundamental and single base for all calculations (Clark, 1976, p.77).[40] Ostwald's *energetics* programme had the aim of "giving a physical interpretation to those purely mathematical operations" (Clark, 1976, p. 77). This will in turn enable one "to connect more tightly from the point of view of physical intuition the premises with the conclusion and so make the deduction more transparent" (Ostwald, 1896, as cited in Clark, 1976, p. 77). In other words, he tried the following:

> Derive the two laws of thermodynamics as consequences of the properties of the fundamental substance energy, the aim of the programme being to divest thermodynamics of its phenomenological character and thus to arrive at new laws, inaccessible by pure thermodynamic techniques. (Clark, 1976, p. 91)

A much more heroic attempt came from Planck (1897). He tried to derive the second law from radiation damping in which his general aim was (still not resolved as far as

[40] Note that Chemero (2009) portrays Boltzmann as a realist and does not mention his anti-realist turn when confronted with the degenerating phase of the atomistic programme. Boltzmann (1900) is explicit about this: "the controversy as to whether matter or energy was the only existing reality appeared to me to be a decided relapse to the old metaphysical point of view which we believed we had overcome, and a violation of the principle that all theoretical concepts are constructive images only" (p. 247).

I know) to open "the prospect of a possible general explanation of irreversible processes by means of conservative forces -- a problem that confronts theoretical physics more urgently every day" (Plank, 1897, as cited in Clark, 1976, p. 91). These two opposite movements showcase an instance where realism and instrumentalism themselves become instruments to improve issues with one's heuristic.

When it comes to Gibsonian ecological psychology and affordances my general attitude is that they should be derivable from more fundamental assumptions and not, by themselves, constitute a part of a heuristic (see §6.3 for more on this). In contrast, Chemero's (2009) heuristic not only has affordances in its hard core but also, he argues for realism by expanding *entity realism* (p. 192). According to this expanded version, if an organism can use affordances in its successful actions and if it can manipulate other things using affordances then these are real. But a distinction is necessary here: one must distinguish between "real effects" and "real constituents." For example, pressure is an entity at the phenomenal level and has real effects but whether it is a constituent of the basic fabric of the universe is a different question. According to contemporary physics, it is not a constituent but only an averaging effect of more basic constituents. But then a methodology obeying Chemero's version of entity realism must treat pressure and atoms just as real. Thus, his proposal stays at the phenomenal level and even worse, it collapses any possible distinctions between levels. In contrast, Ostwald and Planck tried substantial changes to their programme to endow phenomenological thermodynamics with realism. I think such intuitions of first-rate scientists in similar situations should be seriously considered by cognitive scientists worried about the realism aspect of their programmes.

CHAPTER 6

INTERACTIVISM AND ITS APPRAISAL

The appraisals given so far showed that computationalism has the strongest methodology as compared to pure dynamical modelling and radical embodied cognitive science. This chapter will argue that interactivism is at least as strong as computationalism when it comes to heuristic power. I will begin with the relevant portions of interactivism in the first section and review how the key computationalist insights can be captured within interactivism. Then, in the next section, I will first argue that this latter fact does not provide a shortcut for appraising interactivism so that one must make explicit its heuristic (§6.2.1). This will be done in §6.2.2.

6.1 Interactivist representation and the derivation of computationalism

Interactivism gives an action-based account of emergent representation within a process ontology. This ontology in turn incorporates a cascade of emergences beginning from autocatalytic chemical processes to single cell organisms and all the way up to whole persons (with values and ethics) (Bickhard, 1993, 1998a, 2020). Hence the following will be restricted to an overview of interactivist representation and the way it captures the basic properties of computationalist representation within its framework.

The interactivist account of representation is as follows. Confronted with an object (say, a fly flying in front of a frog), the organism must be first causally influenced by it (say, by light, sound, smell, touch, etc.). This is called "contact" (Bickhard, 1993, p. 311). This causal influence will continue *as* a causal influence within the organism (say, through the retinal cells up to the brain). Note that interactions with objects can have multiple time and space scales so that contact is an extended process both in time and space (say, a baby's first contact with a cube toy and the ensuing rotations, shakes, bites, etc.). What are generally taken to be instantaneous perceptions like sight or smell are limiting cases for interactivism. After enough contact the organism will be in a different state than before the contact. That differential between before and after the interaction of the organism is determined *both* by the object or the environment, and the organism itself. Hence that differential can be used to differentiate the type of the object or the environment.

Such a process of differentiation has three essential properties. First, the resolution of the differentiation is necessarily coarse. Since the sole resource for the differentiation is a whole (sub) system organization, many particulars can cause the same system level differentiation (say, any small, moving, black object (flies, pebbles, dots on a screen in a laboratory) can cause the same organizational differential). Hence, only a class of objects are differentiated, and the size of this class depends on the complexity of the organism and the interaction. Second, this differentiation cannot be explicit to the organism, especially if the organizational differential and the interactions involve the whole organism (or close to the whole of the organism). At this stage only an external observer can compare the causal results of the interactions with the type of the objects. Thus, without higher levels of interactions within the organism that are relatively isolated from the consequences of contact, the differentiation is at most implicit. But the organizational difference exists within the organism and so has consequences for further system processes, namely, it is *functional*. Third, there is no representation yet (implicit or explicit) because there is no truth value yet. Everything so far has been purely causal processes.[41]

[41] As we have seen with Chemero's analysis of the Watt governor though, one can also endow such causal processes with representations and then argue that they are explanatorily empty (§2.1).

Truth value emerges in system organizations which can indicate that a particular contact sets up indications of further interaction with the environment that caused the contact.[42] Any such anticipatory system organization must have implicit presuppositions about the context it is in so that the indicated interactions or processes can proceed. Because such processes can *fail* to go as anticipated, truth value emerges and hence representation emerges within such system organizations. The implicit presuppositions of such anticipations constitute the "content" of the representations (Bickhard, 1993, p. 311). To sum up, interactivism's account of representation is future oriented (anticipatory), modal (possible interactions), implicit (at least initially), from general to particular, emergent in system organization, and distinguishes between *contact* and *content*.

Now, these differentiating and indicative system organizations can have structural properties. For example, a cube affords many rotational symmetries. Moreover, this set of symmetries remains the same if the cube is tied to a rope or moved to another room. Such invariances are what objects *are* for interactivism and the resulting structure of all such interaction patterns is called "the situation image." There is implicitness here not only in the above sense but also in that not all possible constructions of the situation image (even for a cube) need to be or can be engaged in explicitly at a given point in time. Such further, potential constructions constitute the "implicit situation image" and the updating of the situation image is called "apperception" (Bickhard, 1998, p. 195).

The above machinery is sufficient to adumbrate the derivation of computationalism within interactivism. Computationalist representations are

[42] Note that although the running example is about external environments, the description of interactions and indications here holds also between subsystems and even subprocesses of an organism as well. This "construction of system modes and manners of processing" for the immediate activity of the system is called "microgenesis" (Bickhard & Campbell, 1996, p. 130).

correspondence type representations and hence their representational content is "encoded" from whatever the representation is in correspondence with. The paradigmatic example of this situation is the Morse code. There are many advantages for using "..." to *stand-in* for "S" such as transmission over long, error prone wires (Bickhard, 1998, p. 202). Similarly, if certain parts of the situation image are encoded into stand-ins that might be useful for the organism in this new form, such as for optimality in learning and off-line processing, then evolution might opt for such constructive capabilities (Bickhard, 1998, p. 199). This is because when a new situation is encountered, although the exact copies (i.e., *not* the stand-ins) of the already constructed organizations of indicators in the situation image might be appropriate content wise, the form of these copies might not fit the new situation. Thus, system organizations that can afford such translators of the relevant parts of the situation image into general stand-ins and back will be advantageous. The products of such translators and the further processing on them will then be like the representations of computationalism and the computational processes on them. Therefore, computationalism can be a possible subfield of interactivism.

Given that interactivism has resources to reconstruct computationalism, the relevant question for this thesis is whether this has any implications when it comes to its methodological appraisal. In the next section I will investigate this issue and give an appraisal of interactivism.

6.2 Appraisal of interactivism

In the following two sub-sections I will first try to give an appraisal of interactivism based solely on its ability to subsume computationalism. Realizing that such a shortcut is not possible in general, I will then (in the next sub-section) give interactivism's appraisal directly by making its heuristic explicit.

6.2.1 An indirect attempt

We have seen in §4 that computationalism has a strong heuristic. We have seen above that the basic structure of computationalist representations and processes on them can be derived within interactivism. Although there are ontological differences between the two, one can still ask if interactivism is methodologically as strong as computationalism? In general, if one type of research programme subsumes another as a subfield or derives another one as a limiting case, what can be said about their heuristic types? If such direct conclusions are possible then lower bounds on the heuristic powers of the subsuming research field can be given without making explicit the heuristic itself.

Unfortunately, such a conclusion does not seem to follow in general. The intuitive appeal of treating the subsuming research programme as better comes from empirical considerations. In such a case it is clear that the more general programme (hereafter: "GP") will not only explain all the facts that the subsumed programme (hereafter: "SP") explains, but it will also explain or predict more. But this covers only half of the criteria presented in this thesis, namely, the first two ad hoc-ness criteria. The criteria of continuity and autonomy are about the nature of the heuristic itself. Since the subsumption is done within the heuristic of the GP, it cannot have implications for directly comparing the two heuristics. If the two heuristics were explicit, then whether and how a direct subsumption (as a subfield or a limiting case) of the heuristic itself of the SP by the heuristic of GP would be possible. Such a task will probably consist of two derivations, namely, one for the hard core of the SP, and one for the positive heuristic of the SP.

This requires making explicit the two heuristics. So, without the explicit heuristic of interactivism, one cannot give an appraisal of it using the appraisal of computationalism. One valuable lesson of this section, though, is that subsumption itself must never be taken as the *sole* measure for methodological superiority. I think this is overemphasized in areas like physics but underemphasized in areas like cognitive science.

6.2.2 A direct attempt

In this section I will give the explicit heuristic of interactivism and argue that it satisfies the autonomous continuous growth requirement. In accord with the way the heuristics of computationalism, and radical embodied cognitive science have been characterized above, the negative heuristic of interactivism can be characterized as follows.

> *Cognition is constituted by the interactions between subsystems within the organism, and between the organism and the environment that result in emergent (interactivism type) representations subject to the principles of interactivism.*

The positive heuristic can be as follows:

I. Make simplifying assumptions about interactions. Construct models of constructions of system organization, and constraints on constructions. The aspects of the system processes and its interactions result in emergent properties that can be taken as a constitutive explanation for the phenomena.

II. These simplifying assumptions should be weakened or replaced to better approximate the real phenomenon under consideration, *e.g.*, a “real” learning task. Here "weakening a simplifying assumption" might mean developing separate process organizations of as yet undifferentiated aspects of a system organization and construction process, e.g., learning and development which are synchronic and diachronic aspects of processes of constructing new system organization (Bickhard & Campbell, 1986, p. 42).

III. Using the specific model created thus far, investigate its location in the cascade of emergences and interactions within interactivism as well as its consequences to derive the known phenomenal properties, e.g., standard representations, folk psychology, language, etc., and predict new ones.

To show that this heuristic satisfies the autonomous continuous growth requirement, I will begin by arguing that it is not ad hoc$_3$. Note that this is built in! Step 3 of the heuristic above explicitly demands continuity by investigating the location of any new model within the whole of interactivism. Bickhard is explicit about such integration when he defines naturalism as "the assumption, or the exploration of the assumption, that there are no intrinsic bounds to inquiry, that further inquiry is always appropriate, and that further integration of understanding is an inevitable consequence of inquiry" (Bickhard, 2020, p. 9). This integration of understanding is to be achieved by proceeding within a framework "in the sense of an *integrated world* [emphasis added] that is not split into distinct metaphysical realms" (Bickhard, 2020, p. 974). Thus, the above heuristic (and step 3 in particular) "yields a progressive refinement, extension, and construction of sub-models within the constraints of more general constraints" (Bickhard, 2020, p. 973). Therefore, the heuristic itself blocks ad hoc_3 stratagems.

What about ad hoc_4 stratagems? Note that the heuristic of interactivism blocks this almost explicitly too. Step 1 in the heuristic demands a constructive approach where the explanandum must be an emergent property so that the model cannot be constructed in a fact dependent manner at least in the way radical embodied cognitive science is fact dependent. The derivation of computationalist representations in §6.1 is typical of interactivism and it is a highly non-fact dependent autonomous model construction. Indeed, an analog of Lakatos's description of Newton's autonomous progress quoted in §3.2 can be given for interactivism. The following is not an exact historical development but that it is *possible* suffices:

Bickhard first worked out his programme for what it means for a simple organism to have an emergent representation in system organization. This required an active agent which produces output so that interaction and anticipation had to take center stage. This in turn made it possible to give an account of simple learning as blind variation and selection in terms of constructive meta-processes on the system processes. But this model *assumed* functional normativity and hence was not satisfying a central principle of interactivism, namely, naturalism understood as integration. Thus, Bickhard moved on to give a model of emergent functional normativity in terms of the asymmetries of the far from equilibrium and near equilibrium thermodynamics. Note that this move was not motivated by any observation that contradicted the existing interactivist theory at that time. At the other end of the emergences, learning by pure random trial and error was forbidden by a touchstone theory, namely, evolution (because of its evolutionary cost). So Bickhard had to come up with system organizations and interactions that allowed the system to interact with its own, more general conditions of uncertainty (e.g., failures to anticipate interactions) such as danger, novelty, or interference. This aspect of a complex enough system organization enabled interactivism to have the surprising conclusion that emotions can be modeled as interactions with conditions of internal uncertainty. The move from the consideration of these restricted meta-processes to more general types of metaprocesses was not only a natural next step in the overall model but also was natural from evolutionary considerations. Thus, a higher-level system that *fully* interacts with this first level system (which is interacting with the external environment) and its meta-processes (learning and emotions) would itself have similar properties. This is because it is also an interacting system organization just like the first system. Possibility of such a second level required and implied the following. First, what is the model of ascent between the levels, and second, iteration of these steps will produce an infinite ascent of levels. At this point Bickhard started to look more anxiously at the facts. One consequence was to correct Piagetian experimenters about the timing of stage transitions because his theory implied that level-3 boundary is half a cycle before as compared to Piaget's formal operations stage.

Note how nicely this fits with an autonomous heuristic's ability to provide a "general outline of how to build the protective belts" right at the start (Lakatos, 1970, p. 137, see also §3.3 above). It is also clear from the above reconstruction that this outline builds the protective belt without ad hoc empirical input from data. Once the models are in place though, the contact with experiments can be established with further differentiations within models, say, whether a domain general ability emerges at age four marking a transition to a level two knowing level (Allen, Çelik, & Bickhard, 2020). Finally, almost any interactivist model generation so far can be seen to follow the above pattern. Hence, interactivism is not ad hoc_4. Since it also subsumes computationalism it is methodologically at least as strong as computationalism.[43]

Note that as opposed to the hard cores of atomists and computationalists, the hard core of interactivism does not seem to have a well-established and well-defined mathematical framework. But the problem here is a practical one and not an in-principle impossibility. One candidate of mathematical formalism for interactivism is *bundle theory* from modern geometry where each fiber over a point in the base space has the structure of possible future interactions (Bickhard & Terveen, p. 323). This converges nicely with Lakatos's observation about strong research programmes that "if the positive heuristic is clearly spelt out, the difficulties of the programme are mathematical rather than empirical" (Lakatos, 1970, p. 136).

Table 4 below is the complete summary of the appraisals given in this thesis, partially ordered in a descending manner. We see that computationalism, interactivism, and atomism has the highest appraisals (none has a “1” in their rows) and are methodologically on a par according to MSRP. Chemero’s Gibsonian radical embodied cognitive science and the pure dynamical approach has the lowest appraisals and are methodologically on a par. Finally, HKB+ and phenomenological thermodynamics are ranked in between these and are methodologically on a par. Note that this thesis only considers continuity and autonomy, so these rankings assume equal empirical progressiveness (i.e., the first two columns in each row have

[43] Note that this conclusion is possible only after first making their heuristics explicit and giving the respective appraisals, and second, showing that interactivism subsumes computationalism. As argued in the previous subsection, one cannot conclude that a methodology is at least as strong another solely based on theoretical subsumption of one by the other.

only zeros). Also note that only considering these two norms is not enough for further ordering within a given rank. For example, among the appraisals of computationalism, interactivism, and atomism, whether a finer grained set of norms can differentiate further methodological differences is not possible to assess at this stage (but see §7.2 for an attempt).

Table 4. The Complete List of Appraisals Given

Appraisal at a non-degenerating phase of a given field	ad hoc$_1$ (theory does not also explain other novel facts)	ad hoc$_2$ (none of the novel facts are corroborated)	ad hoc$_3$ (disconnected series)	ad hoc$_4$ (not autonomous)
Computationalism	0	0	0	0
Interactivism	0	0	0	0
Atomism	0	0	0	0
HKB+ (under a charitable reading)	0	0	0	1
Phenomenological Thermodynamics	0	0	0	1
Gibsonian Radical Embodied Cognitive Science	0	0	1	1
Pure Dynamical	0	0	1	1

6.3 Interactivism and Gibson's ecological psychology

The considerations so far give credence to the view that affordances cannot be a fundamental ingredient in the hard core of a research programme with a strong heuristic. Nevertheless, any such programme must be compatible with affordances and even deduce them. Below are some remarks that interactivism provides one example of such a research programme. I will also point to similarities between

Gibson’s own framework and Chemero in their conceptual maneuvers to move away from standard representationalist theories.

Gibson's affordance approach to cognition, with its emphasis on perceiving objects via the kind of interactions they offer to an organism is close in spirit to interactivism. Historically interactivism was later than Gibson's theory. Thus, as argued by Bickhard and Richie (1983), without a well formulated alternative theory to standard correspondence type representations (such as interactivist type representations), certain measures Gibson took to distance himself from established theories were exaggerated and overstated (p. 25).

This is a general theme in that when confronted with the problems of standard representationalism, a lack of another viable possibility leads to a false dichotomy: There are either representations and processes on them in the standard sense or else, there are no representations. Gibson's move was not getting rid of standard representations but getting rid of internal processes on the representations. He was against the idea that processes on sense data can enhance content. This observation about correspondence type representations, together with a lack of a better alternative, caused him to eliminate such processes and introduce direct pickup. But as Bickhard and Richie (1983) argue, his later theorizing with the introduction of affordances put more and more emphasis on the interactive aspects of his own theory which, indeed, renders direct pick up in the sense of direct encoding impossible from an interactivist perspective (pp. 21-25).

This is a process extended both temporally and spatially so that the whole process, that is, not only the inputs and the outputs of this interaction but also the relations between them, might be needed. The differentiation is achieved via the *pattern* of interaction and not by “any piece or component of the interaction”

(Bickhard & Richie, 1983, p. 14). Direct pick up in the sense of no intermediate representations, no enhancements added to the content by the organism, is however, compatible with interactivism. Note that interactions for direct pick up must involve memory or inferences but only in the *aspectual senses* of these notions because otherwise it will be direct pick up via standard representations (i.e., direct encoding) (Bickhard & Richie, 1983, p. 30). For memory this means that the current pattern of interaction might informationally depend on a past pattern of interaction. Such a memory does not need to be a discrete structure that represents a particular thing but simply be a pointer for switching from one process to another dependent on a previous interaction's outcome (Bickhard & Richie, 1983, p. 23). Thus, they are not memories as an encoding representation of a past event. The account for inferences are similar. Any goal directed system must be able to choose the most appropriate interaction from the available ones. Such a selection can be achieved simply as apperceptive constructions of indicators for further interaction or as control theoretic influence of one process over another (Bickhard & Richie, 1983, p. 27). So, in interactivism, information pick up can be seen as an apperceptive updating of the situation image[44] and as such no intermediate enhancements in the sense that encoding representations and processes on them are needed; it can be direct pick up (Bickhard & Richie, 1983, pp. 24-25).

I have mentioned in §2.1 that Chemero (2009) also presents a similar type of argument against representations (although, unlike Gibson, there is an available alternative model for representations, namely interactivism, that is not subject to his critiques). Moreover, like Gibson, there seems to be a tendency in his theorizing towards putting more emphasis on interactions. Recall that the series of definitions

[44] See §6.1.

he gives for affordances incorporates interactions with affordances 2.0. Nevertheless, because he does not want to postulate internal mechanisms,[45] he opts for abandoning representations all together and postulates affordances as the basic elements of cognition where the mathematical tools of dynamical systems are utilized to model them. In contrast, affordances in interactivism are not the starting point but a natural outcome of the general control theoretic interactive approach.

[45] Note that situation images and apperception form internal organizations and hence they are also forbidden by Chemero's radical embodied cognitive science.

CHAPTER 7

ASPECTS OF MSRP

The previous three chapters showed how MSRP can be used in appraising methodologies in cognitive science. This chapter will change the focus to MSRP itself. In the first section, it is argued that MSRP should be incorporated into the contemporary debates regarding the nature and types of explanation in philosophy of cognitive science. The next section will focus on the notion of autonomy in MSRP and will offer a (very) preliminary account for rendering it more precise and fine-grained, using ideas from interactivism.

7.1 MSRP and the contemporary debates regarding explanation in philosophy of cognitive science

One very important debate in the philosophy of biology and cognitive science concerns mere description versus explanation (Craver 2006; van Gelder 1998; van Leeuwen 2005; Zednik 2011). The intuition behind this distinction is that certain accounts that seem to be explanations might turn out to be mere descriptions. For example, Ptolemy's "explanations" of the movements of the heavenly bodies using epicycles are better taken as descriptions of certain observed regularities. Similarly, modelling a particular cognitive fact via differential equations or some other sophisticated mathematics can also be mere descriptions of the observed data. One aspect of this issue is related to the instrumentalism/realism distinction because confronted with a charge of mere description, one can respond by simply declaring realism. I think van Gelder's nature hypothesis, which we considered in Chapter 4, is one such move. We have also seen that Chemero considers realism as a crucial norm and, because he thinks HKB+ lacks realism, he uses Gibson's ideas as a theory of cognition that has a realist guide to discovery (see Chapter 5 above; see also Chemero, 2000, 2009).

Another aspect of the mere description/explanation distinction is related to phenomenal versus mechanical models (Craver, 2006) which can be seen as a special case of covering-law (Hempel & Oppenheim, 1948) versus mechanical explanations (Bechtel, 2005). The sense in which covering-law explanations are regarded as closer to mere descriptions is given by the received view that covering-law explanations only subsume regularities under general laws without positing underlying mechanisms. Mechanical explanations, however, do posit underlying parts together with their organization. I take it that MSRP's stance regarding these issues are as follows.

As we have seen the unit of appraisal in MSRP is not a single theory but a series of theories. Therefore, taking a particular model or a theory and asking whether it provides a covering-law or mechanical explanation does not make sense from a strict MSRP perspective. One must move up a level and ask: does this heuristic produce a series whose constituents are examples of covering-law or mechanistic models/theories? Newton's research programme produces covering-law theories and for MSRP it has one of the best heuristics around. In contrast, a representationalist computational view of the mind produces mechanistic theories and we have seen that for MSRP it is methodologically as strong as Newton's programme. Hence MSRP is neutral when it comes to favoring covering-law or mechanistic explanations.

I think this is a plus for MSRP because, as Zednik (2011) argues using Thelen, Schöner, Scheier, and Smith's (2001) dynamical account of the A-not-B task and Beer's (2003) dynamical account of categorization, a dynamical model of a cognitive phenomenon does not necessarily imply a covering-law explanation. For example, for the A-not-B task, Zednik (2011) claims that it is an instance of a mechanistic explanation and argues as follows (pp. 248-250). Thelen et al.'s (2001) dynamical field theory model is based on the equation

$$\tau\dot{u}(x,t) = -u(x,t) + S(x,t) + g[u(\acute{x});\ \acute{x}].$$

Here, x is a movement parameter differentiating the positions of A and B, u is the activation level of the *motor planning field* over this region of space, t is time, τ is a temporal decay constant, g is a *cooperativity* parameter, and S is the input field. What is important for us is the following decomposition of S:

$$S(x,t) = S_{task}(x,t) + S_{specific}(x,t) + S_{memory}(x,t).$$

In this decomposition, the first one is the *task input* and models the target locations in the field (e.g., two Gaussian distributions on A and B) (p. 18). $S_{specific}$ models the specific cues given favoring A or B (e.g., "a quick wave of the hand over the target" (p. 19)), and S_{memory} models the bias to the field resulting from the history of previous decisions regarding reaching towards A or B (p. 20).

Zednik (2011) argues that such a decomposition provides a mechanistic explanation because it provides components of a mechanism. Here, mechanism is understood as an "organized activity of a particular collection of simpler processes" (p. 250). Thus, what is important is not whether one uses differential equations, but their form and interpretation. If Zednik's account is accepted as valid, a heuristic might produce a series with mixed elements: for example, the heuristic of the pure dynamical view can generate models which can be examples of covering-law (e.g., HKB) as well as mechanistic (e.g., A-not-B) explanations. Hence, it is good that MSRP's evaluations are robust against such variations.

Concerning the mere description *versus* explanation distinction, however, due to its requirement of autonomous continuous growth, MSRP is biased against heuristics which produce models that are mere descriptions or are merely phenomenological. We have seen that the pure dynamical view of the mind and radical embodied cognitive science do not receive MSRP's highest appraisals. I submit that this should be viewed as an independent (and more general) insight provided by MSRP regarding the distinctions between dynamical approaches and covering-law/mechanistic explanations in the received view of the debate. This can be seen as a convergence with Zednik's arguments mentioned in the previous paragraph.

7.2 Remarks on a possible "interactivisation" of MSRP: Autonomy and being far from equilibrium

The concept of autonomy is vague in Lakatos (1970). Given its importance, a precise formulation should be considered. The following is one such attempt. The motivation for this attempt is not only philosophical clarity. One important consequence might be the ability to differentiate between autonomous research programmes. We have seen that according to MSRP computationalism and interactivism are methodologically on a par because both are autonomous. Perhaps with a more precise formulation of autonomy, these two research programmes might be differentiated within MSRP.

A clarification is needed before moving on. A possible misconception related to the autonomous progress of a research programme is the conflation of anticipating new models with the process of simply working out the "logico-mathematical problem of deriving empirically testable predictions" (Musgrave, 1976, p. 469).

However, even accepting autonomy in this restricted sense of working out the logico-mathematical problem of deriving predictions will suffice for the appraisals given above because the issue concerns the manner in which the steps constituting the derivations for predictions are made. For example, there is a clear difference between the phenomenological thermodynamicists' and atomists' derivations of empirically testable predictions due to the fact dependency of the former. Similar differences hold between the pure dynamical approach and radical embodied cognitive science, on the one hand, and computationalism and interactivism on the other.

Although Musgrave's warning is legitimate, I think his sense of autonomy is too restricted. The proper understanding of an autonomous research programme's empirical dependence should be based on interpreting the role of data as neither the source nor the target of theory generation, but as a modulator of it. Hence the issue is not solely empirical dependence. There are also metaphysical constraints for metaphysical laws. A classic example is special relativity's demand that any fundamental law of physics must be invariant under Lorentz transformations. The existence of such constraints shows that autonomy is not only deriving the empirical consequences from the metaphysical constituents of a heuristic but also the possibility of selection principles within available metaphysical alternatives (Bickhard, 2002).

Two different senses of autonomy are relevant for this thesis. The first one is autonomy as independence from the environment. The formulation and application of autonomy in the previous chapters seem to be close to this sense because the emphasis was on independence from external input during theory generation. However, some of the more fundamental aspects of MSRP seem to be closer to the sense of autonomy characterized by Christensen and Bickhard (2002) as necessarily interactive-self-governance (p. 17). Hence, I will first review their sense of autonomy and then argue why this sense is the relevant one for MSRP.

In Christensen and Bickhard's (2002) formulation of autonomy, interactions with the environment are necessary because they are ontologically constitutive: "a system is autonomous if it interactively generates the conditions required for its existence" (p. 17). There are two aspects here: "firstly, autonomous systems are cohesive in the sense that they interact with the environment as a causally integrated whole; secondly, the conditions required for the cohesion of the system are, at least partly, generated by the system itself" (p. 17). The paradigmatic instantiations of this sense of autonomy are far-from-thermodynamic-equilibrium (hereafter: "FFE") systems that are *self-maintenant*. FFE systems are *open systems* in the sense that there is a continuous exchange of energy and matter with the environment. Self-maintenant systems are a special case of FFE systems because they must interact with the environment to maintain the condition of being in the FFE regime. This is because in equilibrium they cease to exist.

The following are the standard examples for the above differentiations. A thin volume of water heated from below is an FFE system because heat is continuously provided to the system. Moreover, after a threshold temperature convection cells will form a hexagonal tiling of the slab; namely, *Bénard cells* will form. Such self-organizing is a characteristic of FFE systems and is the reason for their importance in understanding living organisms. But note that a Bénard cell organization is not a self-maintaining FFE system because if the heat source is removed, the cells cease to exist. In contrast, a candle flame is a self-maintenant system (though within specific physical constraints – see below). To see this note that in order to exist the temperature of the wick must be kept above a certain threshold which in turn requires source materials as input and waste as output. The crucial distinction from usual system organizations, such as a stone, is that all these maintenance processes to preserve the condition of being in an FFE regime are what a candle flame itself *is*. And unlike the Bénard cells, if you remove the external heat source from the candle, that is if you remove the lighter, the organization of processes does not cease, and the candle continues to burn. But when the wax runs out, the candle flame cannot do anything other than cease to exist. A single cell organism, in contrast, can tumble and swim in a new direction to increase its chance of finding a sugar source. Thus, with living organisms we have a higher degree of autonomous system organization. These systems not only self-maintain being FFE, but they also maintain this property of self-maintenance. Hence, they are called *recursively-self-maintenant systems* (Bickhard, 1993, p. 307).

It is important in what follows to note that the interactivist notion of autonomy is a systemic property that can be instantiated by any system organization satisfying the two conditions in its definition. Christensen and Bickhard (2002) state that they are not defining their concepts in terms of evolutionary process and that the "theory of autonomy can help to *explain* evolutionary units and evolutionary functions, but it is not recursively defined in those terms" (p. 16). Thus, it can be applied to, say, artificial agents or institutions as well as to organic agents or species. In this regard I will try to apply this conceptualization of autonomy to research programmes in general as defined by MSRP. Treating a research programme as an "objectively" interrelated system processes is close in spirit with Lakatos's anti-psychologist rationally reconstruction of standards of growth. For example, he asks "*Can there be any objective (as opposed to socio-psychological) reason to reject a programme, that is, to eliminate its hard core and its programme for constructing* *protective belts?*” (Lakatos, 1970, p. 155). We have seen that his answer is a rival research programme with excess corroborated content. He later reiterates this aspect of MSRP by stating that his “concept of a ‘research programme’ may be construed as an objective, ‘third world’ reconstruction of Kuhn’s socio-psychological concept of paradigm” (Lakatos, 1970, p. 179) where “third world” is the “world of propositions, truth, standards: the world of objective knowledge” (Lakatos, 1970, p. 180). Hence, I will treat research programmes as independent from knowing subjects as much as possible.

Now, the reasons for the relevance of this sense of autonomy to MSRP are the following characteristics of MSRP. The processes that built the protective belt around the hard core of the research programme are like the constitutive interactions of a self-maintenant system. The continuity requirement of MSRP, namely, what a non-ad hoc$_3$ heuristic provides to a research field, is like the cohesion condition of autonomy. Moreover, the selective control of interactions with the environment of a recursively self-maintenant system is very much like Lakatos's description of what a mature autonomous heuristic enables for its practitioner, namely, deciding on what parts of the data to focus attention on and what parts to ignore together with the order of problems to work on (§3.3). This also fits with one of the dimensions of autonomy Christensen and Bickhard (2002) emphasize, namely, "collectively imposed constraints on membership of the system" (p. 22). But to put these intuitions on a sound footing, the relevant identifications of "environment" and "being FFE in a given environment" must be made within MSRP.

Since every research field is operating in a sea of anomalies, it seems that the "environment" can be taken to be just that. But for reasons that will become clear below, the definition of "environment" in MSRP should be taken more general than this. Thus, I define the "environment" in MSRP as an abstract *space of puzzles* which includes not only anomalies, but also facts, anticipated facts, and theoretical puzzles. For my purposes, the finer distinctions between these are not needed because any prediction of a heuristic can be seen as an anomaly until experiments confirm it, any anomaly can be seen as a puzzle to be explained later within a heuristic, any novelty hence anticipation does not depend on objective time, and so on. For example, Newton's gravitation required action at a distance, Maxwell's electromagnetics implied that planetary model of the atom must be wrong, quantum mechanics implies non-locality, precession of Mercury's orbit was a novel prediction of general relativity but was known for a long time, general relativity implied an expanding universe (and Einstein tried to correct it but then it was empirically corroborated), computationalism confronted frame problems, predictive free-energy models created dark room paradoxes, and so on. Therefore, the environment within which research fields dwell can be taken as an abstract and dynamical space of puzzles that includes all these.

Note that this environment is already FFE in the sense that these puzzles are driving the heuristic for generating the series of theories, much like the heat driving the Bénard cells. There are two issues here. First, we have seen that Bénard cells are not a self-maintenant system because they do not contribute to the external heat source. Thus, when it comes to a heuristic, for it to be self-maintenant it must contribute to the space of puzzles. Second, if being FFE is dwelling in a space of puzzles, then an account of equilibrium must be given also.

As for the first issue, recall that a positive heuristic does not merely generate a protective wall around the hard core but must generate further facts, and must have excess content. Lakatos (1970) is clear about such a requirement: "*a given fact is* *explained scientifically only if a new fact is also explained with it*" (p. 34). Since this must be the case for any non-degenerating heuristic in MSRP, it follows that the processes of coming up with a series of theories or models contribute to the puzzle space. Therefore, all such heuristics contribute to their condition of being FFE and hence are self-maintenant.

As for the second issue, if being in the FFE regime is dwelling in a space of puzzles, then equilibrium must be taken as a space where there are no puzzles, or at least no "non-trivial" puzzles. This can be seen as an analog of thermodynamic death which our universe seems to be heading toward and is characterized by the unavailability of "usable" energy that can do work. Now, assume that such a "puzzle death" is attained by a research programme. This means that it not only managed to explain everything in the universe but also is not producing any anomalies or facts of its own. This implies that a research programme ceases to be one if it succeeds in its aims because, as we have seen in the previous paragraph, in MSRP a scientific explanation given for a fact is scientific only if it explains some other new fact. Therefore, the best research programme must end the notion of "scientific explanation." Note that this is a general issue in MSRP that exists independently of the autonomy considerations given in this section. Hence, one way to interpret this is to take it as a reductio and try to change formulations in MSRP itself. Another way to interpret it is to take it as a "proof" that the number of different kinds of puzzles in our universe must be infinite. Perhaps in such a limit case coming up with new puzzles is the sole new fact in the universe. Thus, a heuristic producing such puzzles is contributing to its own conditions of existence in a very strong sense. Yet another way of interpreting is to assume that being a scientific research programme is not an all-or-nothing phenomena but a matter of degree: the more puzzles it solves the less scientific it gets. I do not have knock down arguments to clarify these issues, but they might require an examination at the level of metaphysics. For example, as opposed to a substance metaphysics, a process metaphysics, where organization of processes take center stage, might be more accommodating to some of the above (e.g., the existence of infinitely many puzzles).

Now, the above considerations imply that any theoretically progressive research programme is self-maintenant, and hence autonomous in Christensen and Bickhard's (2002) sense because each element in such a series of models has excess content (§3.1). Conversely, a degenerating research programme, either because of ad hoc_3 stratagems or because of a lack of excessive content, is not autonomous. This means that the only way to differentiate non-ad hoc_4 research programmes from ad hoc_4 research programmes is to show that they instantiate recursively self-maintenant properties. That is, their constitutive activities should maintain the activity of coming up with excess content (i.e., new puzzles). But note that the strong heuristics we have seen above, say atomism or interactivism, do exactly that. When confronted with a "hard" puzzle, they either find new paths or create connected and extending niches within the puzzle space. A weak heuristic, in contrast, simply "halts" or degenerates. For example, Newton created and solved numerous puzzles until he reached a model with realistic, bulging planets affecting each other, and so on.

However, I think there is a serious equivocation in the attempt to equate the property of recursive self-maintenance with non-ad hoc_4-ness. This can be best described via making a direct comparison with the candle flame. The FFE condition of heat corresponds to the FFE condition of dwelling in the puzzle space. When the wax begins to run out, a (fictitious) recursively self-maintenant candle flame must find a new wax source. Such a capacity corresponds to the external input independent stratagems of a non-ad hoc$_4$ research programme. But so far in my account there is only an undifferentiated abstract puzzle space. So, both heat and wax (and oxygen, and environment, etc.) correspond to the same kind of entity, namely, "puzzles." Thus, a differentiation must be made here. Moreover, and more importantly, postulating an "abstract puzzle space" can only be a temporary place holder in an account which considers naturalism (i.e., the regulative ideal of eliminating all ad hoc-ness – see §6.2.2) as one of the important criteria for model building. In other words, a model for the "abstract puzzle space" must be given.

For such a task the *knowing levels model* within interactivism can be a source of modelling ideas (Campbell & Bickhard, 1986). Now, the definition of "knowing" in interactivism is not restricted to agents with higher cognitive capacities. "To know something" is to successfully interact with that thing according to anticipations. For example, a single cell organism swimming up a sugar gradient with the aim of reaching an environment with higher sugar content "knows" about such gradients. The "levels" in this model are defined in a recursive manner. The first level is constituted by those processes of a system organization that interacts with the environment directly. These interactions include self-maintaining and recursively-self-maintaining interactions with the environment. The second level is those processes of system organization that interact with the first level system processes. The third, fourth, and all the higher levels are defined similarly. Naturally, higher abilities and abstractive powers require higher knowing levels. For example, the ability to formalize logical forms such as logical implication requires a level four knowing in this model (Campbell & Bickhard, 1986, p. 108).

Now, the needed differentiations for the abstract puzzle space can be given using these levels. The empirical puzzles and the empirical input, which were seen as obstacles to autonomy in this thesis,[46] can be taken as belonging to the environment that the first level interacts with. Idealizing assumptions and metaphysical elements of the heuristic belong to higher level "environments" where higher knowing levels are interacting with these environments. Note that for this model to work, it should be possible to be in the FFE regime at each level. For example, level three can be a (recursively) self-maintaining system relative to level two which *is* its environment. Then, the difference between an ad hoc_4 and non-ad hoc_4 research programme can be given in terms of the relative autonomy of the higher levels as compared to the first level. A research field with only level 1 interactions (say, pure dynamical stance) will thus be totally fact dependent.

Unfortunately, at this stage I cannot provide more than the above considerations which are only a rough sketch of ideas. Assuming that possible dead ends and inconsistencies can be dealt with, knowing levels model and interactivist understanding of autonomy can help discern many different types of heuristics in MSRP. Thus, a differentiation between non-ad hoc_4 research programmes can then become possible.

[46] And also lead to a notion of autonomy which is more compatible with autonomy as independence from the environment.

CHAPTER 8

CONCLUSION AND FURTHER PROSPECTS

This thesis has sought to show that MSRP is quite relevant to current methodological issues in cognitive science. With its non-local tools of appraisal and emphasis on metaphysics, MSRP can provide better resolution for comparing various methodologies (as we have seen with Chemero’s case). It can also supplement methodological debates such as that between mechanical *versus* covering-law explanations. These are strong enough merits to consider MSRP as at least one essential tool for any methodological considerations in cognitive science.

Although I do not agree with Chemero's norms, I agree with many aspects of his criticism regarding teleological and correspondence type representations, the computational theory of mind, and his highlighting careless use of dynamical systems. The reason for defending computationalism in this thesis was solely for its methodological properties and strength. As we have seen, interactivism, in contrast, not only has a strong appraisal from MSRP but also can subsume computationalism via deriving its type of representations and by explaining how any agent can ever achieve the kinds of representational states computationalism assumes.

I would like to finish with possible future directions of research stimulated by the issues discussed in this thesis. First, further work in philosophy of science which criticized and improved Lakatosian insights should be pursued. Although metaphysics is important for Lakatos (e.g., the hard core), his criteria for progress is still ultimately empirical progress. I think that there should be some kind of conceptualization of criteria for "metaphysical progress." Laudan's (1977) work on *research traditions* and Bickhard's (2002) work on *critical principles* seems to be good starting points.

Second, and related to the first issue above, formalizations of cohesion and autonomy should be investigated so that whether MSRP can differentiate between continuous and autonomous programmes can be investigated. One aspect of this can consist of further work in looking for ways to apply existing notions of autonomy to MSRP. Another aspect can be trying to develop custom made notions of autonomy inspired directly by autonomous research programmes.

Third, I progressively came to the realization that interactivism has enough ontological and metaphysical resources to initiate a strictly interactivist philosophy of science. This was a result of trying to use the interactivist notion of autonomy, but I think the "knowing levels" model, and critical principles (mentioned above) for rationality are aspects with immediate consequences. Although Bickhard (2020) has been aware of such a potential and provides general outlines, a thorough engagement with interactivism solely for the sake of philosophy of science looks very promising.

Finally, I would like to conclude with Lakatos's thoughts on MSRP and heuristic advice. One misconception related to MSRP's appraisals is whether an appraisal that reveals weaknesses of a particular methodology should mean abandoning it altogether. For example, we have seen that Chemero’s defense of anti-representationalism was not as strong as he claims it to be. Does this mean that cognitive scientists sympathetic to Chemero’s programme should switch to the stronger computationalist programme? Lakatos (1976) is explicit about this possible misreading of his philosophy of science :

> One may rationally stick to a degenerating programme until it is overtaken by a rival and *even after*. What one must *not* do is to deny its poor public record. Both Feyerabend and Kuhn conflate *methodological* appraisal of a programme with firm *heuristic* advice about what to do. It is perfectly

> rational to play a risky game: what is irrational is to deceive oneself about the risk. (pp. 15-16)

The main reason for refraining from giving heuristic advice is that a degenerating series might turn into a progressive one and *vice versa*. This fallible nature of the appraisals might be seen as a weakness of MSRP. Unlike his critique of the anticipatory aspect of autonomy, Musgrave (1976) firmly supports MSRP in this issue by distinguishing between giving advice to an individual versus to a community. After such a “Lakatosian move,” (that is, changing the unit of the evaluation) Musgrave argues that MSRP does indeed have the resources for heuristic advice (p. 480): Confronted with two rival research programmes, one of which, say, has more types of ad hoc-ness, the scientific community should devote *most* of its resources to the stronger programme. I take it that this also means that the duty of the weaker programme is to become aware of the “risks” and not be content with “business as usual.” In this sense Chemero’s attempts to amend the pure dynamical view are quite valuable and should be taken into serious consideration.

REFERENCES

Abraham, R. H., & Shaw, C. D. (1987). Dynamics: A visual introduction. In F. E. Yates (Ed.), *Self-organizing systems: The emergence of order* (pp. 543-598). New York and London: Plenum Press.

Aerts, D. (2007). *Quantum interference and superposition in cognition: Development of a theory for the disjunction of concepts*. arXiv. http://arxiv.org/abs/0705.0975

Allen, J. W. P., Çelik, B., & Bickhard, M. H. (2020). Age 4 transitions: Reflection as a domain-general development for explicit reasoning [Manuscript submitted for publication].

Asarin, E., & Maler, O. (1994). On some relations between dynamical systems and transition systems. In S. Abiteboul & E. Shamir (Eds.), *Automata, Languages and Programming. ICALP 1994. Lecture Notes in Computer Science, 820* (pp. 59-72). Jerusalem, Israel, July 11-14, 1994. Berlin, Heidelberg: Springer.

Bechtel, W. (2005). Explanation: A mechanist alternative. *Studies in History and Philosophy of Biological and Biomedical Sciences*, *36*, pp. 421–41.

Beer, R. D. (2003). The Dynamics of active categorical perception in an evolved model agent. *Adaptive Behavior*, *11*(4), pp. 209-43.

Beltrami, E. (1987). *Mathematics for dynamic modeling*. Boston: Academic.

Bickhard, M. H. (1980). *Cognition, convention, and communication*. New York: Praeger Publishers.

Bickhard, M. H. (1993). Representational content in humans and machines. *Journal of Experimental and Theoretical Artificial Intelligence*, *5*, pp. 285-333.

Bickhard, M. H. (1997). Emergence of representation in autonomous agents. *Cybernetics and Systems*, *28*(6), pp.489-498.

Bickhard, M. H. (1998a). Levels of representationality. *Journal of Experimental and Theoretical Artificial Intelligence*, *10*, pp. 179-215.

Bickhard, M. H. (1998b). Robots and representations. In R. Pfeifer, B. Blumberg, J. A. Meyer, & S. W. Wilson (Eds.), *From Animals to Animats 5. Proceedings of the Fifth International Conference on Simulation of Adaptive Behavior* (pp. 58-63). Zurich, Switzerland, August 17-21, 1998. Cambridge, Mass: MIT Press.

Bickhard, M. H. (2002). Critical principles: on the negative side of rationality. *New Ideas in Psychology*, *20*, pp. 1-34.

Bickhard, M. H. (2015a). Toward a model of functional brain processes I: Central nervous system functional micro-architecture. *Axiomathes*, *25*, pp. 217-238.

Bickhard, M. H. (2015b). Toward a model of functional brain processes II: Central nervous system functional macro-architecture. *Axiomathes*, *25*, pp. 377-407.

Bickhard, M. H. (2016). The anticipatory brain: Two approaches. In V. C. Müller (Ed.), *Fundamental Issues of Artificial Intelligence* (pp. 259-281). New York: Springer.

Bickhard, M. H. (2020). *The whole person* [Manuscript in preparation].

Bickhard, M., & Campbell, R.L. (1996). Topologies of learning and development. *New Ideas in Psychology*, *14*(2), pp. 111-156.

Bickhard, M. H., & Richie, D. M. (1983). *On the nature of representation: A case study of James Gibson's theory of perception.* New York: Praeger Publishers.

Bickhard, M. H., & Terveen, L. (1995). *Foundational issues in artificial intelligence and cognitive science*. Amsterdam: Elsevier.

Block, N., & Fodor, J. (1972). What psychological states are not. *The Philosophical Review*, *81*, pp. 159–181.

Boltzmann, L. (1900). The recent development of method in theoretical physics. *Monist*, *11*, pp. 226–57.

Bournez, O., & Cosnard, M. (1996). On the computational power of dynamical systems and hybrid systems. *Theoretical Computer Science*, *168*(2), pp. 417-59.

Campbell, R. L., & Bickhard, M. H. (1986). *Knowing levels and developmental stages*. Basel: Karger.

Chemero, A. (2009). *Radical embodied cognitive science*. Cambridge: The MIT Press.

Chemero, A. (2000). Anti-representationalism and the dynamical stance. *Philosophy of Science, 67*, pp. 625-47.

Christensen, W. D., & Bickhard, M. H. (2002). The process dynamics of normative function. *The Monist*, *85*(1), pp. 3-28.

Clark, A. (2001). *Mindware*. Oxford: Oxford University Press.

Clark, P. (1976). Atomism versus thermodynamics. In C. Howson (Ed.), *Method and Appraisal in the Physical Sciences* (pp. 41-105). New York: Cambridge University Press.

Craver, C. (2006). When mechanistic models explain. *Synthese, 153*, pp. 355-76.

Dretske, F. (1981). *Knowledge and the flow of information*. Cambridge, Mass: MIT Press.

Eliasmith, C. (1996). The third contender: A critical examination of the dynamicist theory of cognition. *Philosophical Psychology, 9*(4), pp. 441-63.

Eliasmith, C. (2003). Moving beyond metaphors: Understanding the mind for what it is. *The Journal of Philosophy*, *C*(10), pp. 493-520.

Fodor, J. A. (1975). *The language of thought*. New York: Thomas Y. Crowell Company.

Fodor, J. A. (1981). *Representations*. Cambridge, Mass: MIT Press.

Fodor, J. A. (1990). *A theory of content and other essays*. Cambridge, Mass: MIT Press.

Giunti, M. (1991). *Computers, dynamical systems, phenomena, and the mind* Unpublished PhD thesis). Indiana University, Indiana, USA.

Grush, R. (1997). The architecture of representation. *Philosophical Psychology*, *10*, pp. 5-24.

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

Hempel, C., & Oppenheim, P. (1948). Studies in the logic of explanation. *Philosophy of Science*, *15*(2), pp.135-175.

Kaipainen, M., & Port, R. (unpublished manuscript). Speech cycling.

Lakatos, I. (1970). Falsification and the methodology of scientific research programmes. In I. Lakatos & A. Musgrave (Eds.), *Criticism and the Growth of Knowledge* (pp. 91-195). London: Cambridge University Press.

Lakatos, I. (1976). History of science and its rational reconstructions. In C. Howson (Ed.), *Method and Appraisal in the Physical Sciences* (pp. 1-39). New York: Cambridge University Press.

Lakatos, I. (1978). Science and pseudoscience. In J. Worrall & G. Currie (Eds.), *Philosophical Papers, vol. 1: The Methodology of Scientific Research Programmes* (pp. 1-7). Cambridge: Cambridge University Press.

Lakatos, I. & Zahar, E. (1978). Why did Copernicus's research programme supersede Ptolemy's. In J. Worrall & G. Currie (Eds.), *The Methodology of Scientific Research Programmes* (pp. 168-192). New York: Cambridge University Press.

Laudan, L. (1977). *Progress and its problems*. Berkeley: University of California Press.

Maxwell, J. C. (1876). Atom. *Scientific Papers*, *2*, pp. 445-484.

Millikan, R. (1984). *Language, thought, and other biological categories*. Cambridge, Mass: MIT Press.

Millikan, R. (1993). *White queen psychology and other essays for Alice*. Cambridge, Mass: MIT Press.

Musgrave, A. (1976). Method or madness? In R. S. Cohen, D. Davidson, G. Nuchelmans, & W. C. Salmon (Eds.), *Essays in Memory of Imre Lakatos* (pp. 457-91). Boston: D. Reidel Publishing Company.

Ostwald, W. (1896). Zur Energetik. *Annalen der Physik und Chemie*, *58*, pp. 154-167.

Planck, M. (1945). *Treatise on thermodynamics* (A. Ogg, Trans.) New York: Dover (Original work published 1897).

Planck, M. (1897). Über irreversible Strahlungsvorgänge. Erste Mitteilung. *Physikalische Abhandlungen und Vorträge*, *1*, pp. 493-504.

Rescorla, M. (2020). The computational theory of mind. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*. Retrieved December 18, 2020, from https://plato.stanford.edu/

Smith, B. C. (1996). *On the origin of objects*. Cambridge, Mass: MIT Press.

Thelen, E., Schöner, G., Scheier, C., & Smith, L.B. (2001). The dynamics of embodiment: A field theory of infant preservative reaching. *Behavioral and Brain Science*, *24*, pp. 1-34.

Thelen, E., & Smith, L.B. (1994). *A dynamic systems approach to the development of cognition and action*. Cambridge, MA: The MIT Press.

Thrun, S., Burgard, W., & Fox, D. (2006). *Probabilistic robotics*. Cambridge, MA: MIT Press.

Turing, A. (1936). On computable numbers, with an application to the Entscheidungs problem. *Proceeding of the London Mathematical Society*, *42*, pp. 230-265.

van Gelder, T. (1995). What might cognition be, if not computation? *Journal of Philosophy*, *91*, pp. 345-381.

van Gelder, T. (1998). The dynamical hypothesis in cognitive science. *Behavioral and Brain Sciences*, *21*, pp. 615-665.

van Gelder, T. (1999). *Revisiting the dynamical hypothesis*. CiteSeerX. https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.35.5708&rep=rep1&type=pdf

van Leeuwen, M. (2005). Questions for the dynamicist: The use of dynamical systems theory in the philosophy of cognition. *Minds and Machines*, *15*, pp. 271-333.

Volkenstein, M. V. (2009). *Entropy and information* (A. Shenitzer & R. G. Burns, Trans.). Berlin: Birkhaüser. (Original work published 1986)

Zahar, E. (1976). Why did Einstein's programme supersede Lorentz’s? In C. Howson (Ed.), *Method and Appraisal in the Physical Sciences* (pp. 221-275). New York: Cambridge University Press.

Zednik, C. (2011). The nature of dynamical explanation. *Philosophy of Science*, *78*(2), pp. 238-63.