IMPLICIT DEFINITION AS UNIFYING LOGIC OF MENTAL REPRESENTATION

NEZİHE MÜGE KUYUMCUOĞLU TÜTÜNCÜOĞLU

BOĞAZİÇİ UNIVERSITY

2023

IMPLICIT DEFINITION AS UNIFYING LOGIC OF MENTAL REPRESENTATION

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

Nezihe Müge Kuyumcuoğlu Tütüncüoğlu

Boğaziçi University

2023

DECLARATION OF ORIGINALITY

I, Nezihe Müge Kuyumcuoğlu Tütüncüoğlu, 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……

Date ……

ABSTRACT

Implicit Definition as Unifying Logic of Mental Representation

The problem of intentional nature of mental representations remains a priority in analytic philosophy and cognitive science. Its impact on the foundations of logical and ontological statements, such as variable referencing over objects, reveals its fundamental importance. The prevalent notion of mental representation, influenced by external representation, treats representations as static stand-ins. This poses challenges in accounting for dynamic aspects such as mensurable change and unity of representations, misrepresentation, and errors detectable by agents. Understanding mental representations therefore has implications for our best understanding of knowledge and learning. This thesis investigates the interactivist notion of mental representation, modeled after implicit defining, as an alternative. In this view, mental representations do not strictly correspond to their contents but rather "define" them, analogous to how axiomatic systems define their terms. The concept of implicit representing sheds light on the contentful nature of habits as a foundational mode of interaction. The interactivist alternative is explored with a view towards the historical development of analytic philosophy, elucidating the relationship between implicit defining, the logical operation, and the ontology of mental representation. It explicates key concepts such as habit, diagrammatical reasoning, abstraction vs. exemplification, problem of induction, functionalism, inferential role semantics, interventionist theory of causation, measure theory, concepts versus properties, framing, systematicity, and productivity. The thesis traces the reactions of analytic philosophy against pragmatism, offering a fully functionalist, operation-based, and pragmatist perspective that abandons structural

notions. It shifts the core of explanation from linear mechanistic models to emergentist models of constraints and an interventionist account of causality.

ÖZET

Zihinsel Temsillerin Bütünleştirici Mantığı olarak Örtük Tanımlama

Gerek analitik gerekse bilişsel bilim alanlarında, zihinsel temsillerin doğası problemi önceliğini sürdürmektedir. Zihinsel temsil problemine getireceğimiz açıklama, değişkenlerin nesnelere nasıl atıfta bulunabildiği sorusunda olduğu gibi ontolojik önerme yapmanın temel kabul ve ön koşullarını değiştirmektedir. Bu sorulara vereceğimiz cevap, ister dilin kullanılma kuralları aracılığıyla bir çeşit normatif zorunluluk ve olasılık dünyasına işaret etsin, isterse de bir çeşit belirli "psiko-dilsel" dolayım aracılığıyla gerçekleştiği önerilsin, bu nevi bir açıklama getirmeden, en basit mantıksal ve ontolojik formalizasyonlar temelden yoksun kalacaktır. Dolayısıyla zihinsel temsilleri anlamak, bilgi ve öğrenme kuramlarımızı da değiştirecektir. Literatürde yerleşmiş zihinsel temsil kavramı, yazı, resim gibi dışsal temsilden etkilenerek, zihinsel temsilleri de statik ve sabit içeriklere bağlı olarak ele alır. Bu, dinamik yönlerin, gelişimin, yanlış temsillerin ve kişilerin kendi kendilerinin tespit edebileceği hataların açıklanmasında zorluklar ortaya çıkarır. Alternatif bir bakış açısı ise etkileşimci zihinsel temsil teorisidir, bunun temeli ise örtük tanımlama işleminden gelmektedir. Bu görüşe göre, zihinsel temsiller içerikleriyle izomorfik olarak uyuşmaz, ancak onları benzer şekilde "tanımlarlar", aksiyomatik sistemlerin terimleri tanımlaması gibi. Zımni (örtük) temsil kavramı, etkileşim temelli bir temsil biçimi olarak alışkanlıkların içerikli doğasını aydınlatır. Bu tez, analitik felsefe tarihi bakımından etkileşimci alternatifin, örtük tanımlama işleminin ve zihinsel temsillerin ontolojisi arasındaki ilişkiyi açıklar. Alışkanlık, diyagramsal çıkarım, soyutlama vs. somutlaştırma, indüksiyon sorunu, işlevselcilik ve çıkarımsal rol semantiği, müdahaleci nedensellik

teorisi, ölçüm teorisi, kavramlara karşın özellikler, çerçeveleme ve çerçeve sorunları, dilsel sistemlerin sistemliliği ve üretkenliği gibi kavramları ele almaktadır. Bu tezde yapmak istediğim, formalist analitik felsefe ve pragmatist felsefe yaklaşımları arasındaki fay hattını, zihinsel temsil problemi bağlamında ele almaktır. Bu sorunun cevabı son yıllarda felsefe literatüründe zihin felsefesi ve bilişsel bilimin yükselişinin sebeplerini anlamamıza olduğu kadar, aynı zamanda analitik felsefe tarihini de pragmatizme karşı geliştirdiği refleksler bağlamında daha geniş açıdan okuyabilmemizi sağlamak yolunda katkı sağlayacaktır. Zihinsel temsilin anlaşılmasındaki bu temel değişim, yapısal kavramlardan vazgeçerek tamamen işlevsel, işlem temelli ve pragmatist bir perspektife doğru kaymayı başlatır. Bu yaklaşım, açıklamanın odağını yapısal tanımların lineer mekanistik modellerinden, etkileşimli sistemlerin kısıtlamacı ve müdahaleci nedensellik anlayışlarına kaydırır.

CURRICULUM VITAE

NAME: Nezihe Müge Kuyumcuoğlu Tütüncüoğlu

DEGREES AWARDED

PhD in Philosophy, 2023, Boğaziçi University

MA in Philosophy, 2015, Boğaziçi University

BSc. in Management Information Systems, 2006, Boğaziçi University

AREAS OF SPECIAL INTEREST

Philosophy of mind, Philosophy of cognitive science, interactivism, philosophy of psychology

PROFESSIONAL EXPERIENCE

Research Assistant, TÜBİTAK 2232 Program, Boğaziçi University, 2019 – 2022

Reporting User Interface Designer, Pazarlama ve Iletisim Hizmetleri, Koç Holding, 2012 - 2016

Senior Technology Consultant, Accenture Consulting Türkiye, 2010 – 2012

Project Manager, Hewlett Packard Software Consulting Türkiye, 2006 – 2010

AWARDS AND HONORS

Tübitak BİDEB 2250 Graduate Performance Award, August 2022

Honors list, Boğaziçi University, 2006

GRANTS AND SCHOLARSHIPS

TÜBİTAK 2232 PhD Scholarship, 2019 – 2022

PUBLICATIONS

Masters Thesis

Tütüncüoğlu, N. M. K. (2015) A Philosophical Interpretation of the Unconscious in light of Wilfrid Sellars' Reasons – Causes Dichotomy (Unpublished master's thesis). Boğaziçi University, Istanbul, Turkey.

PRESENTATIONS AND TALKS

Tütüncüoğlu, N. M. K. (2023) Representing in Skilled Action. In *Dynamics of Agency: Spontaneity, Automaticity and Change Workshop,* University of Helsinki, Helsinki, Finland.

Tütüncüoğlu, N. M. K. (2020) Framing and Interactivism. In *Interactivism in Perspective Conference*, held online, https://www.ecointeractivism.com/conference.

Tütüncüoğlu, N. M. K. (2019) Relevance Problem as it Applies to Learning from Experience. In Graduate Student Philosophy Workshop, Boğaziçi University, Istanbul, Turkey.

ACKNOWLEDGEMENTS

This thesis has allowed me to work with and learn from wonderful people that I deeply admire. I am grateful to you for helping me to achieve my dream of doing philosophy in my own way, and to become the person I am today: Stephen Voss, Lucas Thorpe, Mark Bickhard, İlhan İnan, and Frank Zenker. I am inspired by your examples and will try to imitate you in the best possible ways.

I would also like to acknowledge my past teachers in Boğaziçi University Philosophy Department, who have taught and supported me in every possible way: Sun Demirli, Johannes Fritsche, and Ken Westphal.

I would also like to mention two friends who have become comrades to me during this period: Seçil Aracı and Oğuz Erdin.

Finally, I would like to thank my family, my dear partner Galip, my dear father Umur, my darling sister Aybüke and my beloved babies Lale and Hayal, for their infinite love and encouragement.

I also acknowledge that I have been supported by TÜBİTAK as a PhD scholar as part of the 2232 project titled: Models, Theories and Research Programmes, 118C257 during my studies.

This thesis is dedicated to my two pillars, Galip and Umur.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

THE PROBLEM OF REPRESENTATION IN LEARNING FROM EXPERIENCE

A key problem for our understanding of "learning from experience" is an understanding of how mental representations enable learning. The link between the content of our mental representations and learning is still not clear in our most recent accounts of cognition. Learning is a process whereby the agent starts from a place of not knowing, and proceeds to acquire new understanding, knowledge, and skills. Mental representations are truth bearing epistemic aspects of the cognitive system that are used by the same system to accomplish cognitive and physical acts. Increased success and accuracy in action can be taken as an existence proof of gain in accuracy and precision of mental representations. Mental representations track, as well as aid in, learning and development of skill. But how specifically mental representations aid learning and what representations are so that they change (or improve i.e., through repetitive practice) is not clear. What kind of things mental representations are, that they are fluid and flexible to change, and yet reliable and stable to guide action and problem solving? [1]

---

[1] In fact, this question can be pushed so far, that we can question the possibility of learning anything new. If pre-existing and pre-determinate knowledge aids learning, (or know-how aids action), how does one go beyond what one already knows? How are genuine inventions possible? The problem has to do with the generation of "representational primitives". There must be a way to account for the primitives, because combining or copying operations already presuppose primitive representational atoms or "units of thought", out of which combinations or from which copying can occur. What is at issue is the origin of the representational primitives. (Allen & Bickhard, 2011; Carey, 2004, 2009)
Fodor (2005) formulates a similar problem with regards to the origin of concepts, but the conclusion he reaches as a result of the paradox must be taken as a reductio ad absurdum: that the acquisition of concepts is impossible – all primitives must be natively present. Here is his formulation of the problem:
"How is one to suppose that a mind can tell a C from a D unless it is already able to think about Cs and Ds? Likewise, how is one to suppose that a mind can infer from Cs being F to Cs being G unless it is already able to think about Cs (to say nothing of Fs and Gs)?" As a result of this reasoning, Fodor

Part of the problem is with a static understanding of mental representations. The literature on mental representation takes the phenomenon of mental representation as full-fledged epistemic semantic entities in the head as a starting point and attempts to give a sufficient explanation of either their truth conditions, or their material/implementational make-up, without a view towards their early origin or genesis.[2] The idea is that of a paradigmatic representation as a "stand-in". This prevalent

---

concludes that all concepts must be innate, and all intellectual invention is a matter of combinations of representations that already exist. This would lead to the conclusion that there is no genuine learning and development.

Since learning is an actual phenomenon, the paradox must be a conceptual one. In many accounts "what is known" connotes something that has already happened, static, timeless, well defined, and fully determinate. In considering mental representations, we must be able to include partial knowledge that aids going beyond what is fully known, so one can still go beyond what is known, albeit unknowingly (D. T. Campbell, 1987, p. 57). In what way knowledge can be partial and how representations enable learning is directly tied to the actual possibility of learning and discovery.

Piaget (1971) links the problem of the origins of knowledge in experience (what he calls genetic epistemology) with the possibility of learning and ultimately with construction as intellectual creation: "…why I consider the main problem of [[genetic epistemology]] to be the explanation of the construction of novelties in the development of knowledge. From the empiricist point of view, a "discovery" is new for the person who makes it, but what is discovered was already in existence in external reality and there is therefore no construction of new realities. The nativist or apriorist maintains that the forms of knowledge are predetermined inside the subject and thus again, strictly speaking, there can be no novelty. In contrast, for the genetic epistemologist, knowledge results from continuous construction, since in each act of understanding, some degree of invention is involved; in development, the passage from one stage to the next is always characterized by the formation of new structures which did not exist before, either in the external world or in the subject's mind. The central problem of [[genetic epistemology]] concerns the mechanism of this construction of novelties which creates the need for the explanatory factors…A great deal of work remains to be done in order to clarify this fundamental process of intellectual creation, which is found at all the levels of cognition, from those of earliest childhood to those culminating in the most remarkable of scientific inventions." (1971, p. 77)

[2] Making the context of justification the focus of philosophy is a positivistic and empiricist tendency, stemming from Carnap and later from Reichenbach arguing against Peirce's abduction. The distinction assumes that the context of justification is independent from the context of discovery, the former being objective and the proper subject of philosophy and the latter being subjective and the proper subject of empirical psychology only. (Carnap, 1967; Glymour & Eberhardt, 2022; Uebel, 2016)

A parallel dispute is the Hume originated problem of the legitimacy of habit, the source versus validity of predictions. Goodman criticizes this formulation of this problem as illegitimate: "I merely want to record a protest against the prevalent notion that the problem of justifying induction, when it is so sharply dissociated from the problem of describing how induction takes place, can fairly be called Hume's problem." (Goodman, 1983, p. 61)

When we talk about justification separately from discovery, we are smuggling in assumptions regarding the origin of knowledge, learning, concepts, which may in fact undermine our accounts of justification to the extent that justification is a dynamic process. The underlying metaphysical assumption is not only that discovery is either blind or heuristical, but also that the discovering agent becomes in

notion of mental representation inherits from the notion of an external representation, a typical material symbol, which is a code that acts as a stand-in for a perceived content. This idea is not only a rhetorical step, merely to give the necessary and sufficient conditions of the phenomena to be defined, it also betrays underlying metaphysical assumptions with received understanding of the nature of mental representations, which is not often scrutinized. The metaphysical assumption is that, if there is such a thing as mental representation (epistemic, truth bearing,[3] informationally rich), it must have the ontological and logical characteristic of a stand-in, that "represents" what (properties of the environment) it stands-in for, like a signpost. This intuitive idea is very foundational to the very conception of a mental representation. But it is wrong.

For instance, as rehearsing improves the coupling between an agent's skilled action and her environment, the covariance between the agent's representations and the actual properties in the environment should increase. In other words, it is expectable that her representations gain in precision. However, how this gain is realized remains unclear. Most commonly, some sort of fidelity in some sort of correspondence is presupposed. However, since representations are typically thought to be tied to contents to be encoded, representations are static (i.e., already fixed). If a representation is static, however, then accounting for misrepresentations (erroneous representations) and for errors that agents themselves can detect (to improve) is already difficult. Add to this the

---

Campbell's words "a passive induction machine accumulating contingencies" (D. T. Campbell, 1960, p. 389).

[3] "Truth-bearing" is used to indicate the semantically evaluable nature of representations. Mental representations can be true or false, correct or incorrect, in their attributions to particular circumstances.

challenge of accounting for changes in representations that occur as an agent's skills improve, a challenge that received views on mental representation arguably cannot meet.

Indeed, criticism of epistemic correspondence, with or without reference to mental representations (or misrepresentations) is ubiquitous in the history of philosophy and cognitive science. There are many "named" problems that point to inconsistencies with prevalent accounts of representations.[4] There are even accounts that give up on the phenomenon of mental representation altogether, for example in Brooks (1991), see Kirsh (1991) for a critique. Regardless of the critiques or on pain of resignation on accounting for the phenomenon of representation, the received characterization of mental representations prevails, so that any alternative models are inconceivable.

There is, however, an alternative account, the interactivist model of representation (and cognition), that suggests that the "stand-in" aspect is not necessary to mental representation (Bickhard, 2009b, 2009a, 2023). That mental representation in its normative, functional, truth-bearing, action-guiding guise can still emerge, and in fact must emerge, without being a stand-in for environmental properties in the first place. And stand-in forms of representation can be derivatively construed from these fundamental representations. If the logical form of a (derivative) stand-in (representation) is that of an explicit definition, then the logical form of a primitive mental representation should be depicted as a "Hilbertian implicit definition" (Freudenthal, 1966; Hilbert, 1902; Resnik, 1974).[5] According to the notion of implicit

---

[4] Symbol grounding problem (Harnad, 1990), Chinese room problem (Searle, 1984, p. 32), epistemic veil (veil of perception) problem, homunculus problem, problem of misrepresentation, problem of intrinsic intentionality, accounting for misrepresentation and others, some of which to be explicated below.

[5] Here I am using "Hilbertian" in a broader than formal sense. As Bickhard does, I take implicit definition in its more general form to apply to presuppositions of interactions, presuppositions that are intrinsically

defining, mental representations do not refer to their contents or strictly correspond to them; they merely "define" them in ways similar to how an axiomatic system "defines" the terms satisfying them. The axioms of geometry are taken as definitions in disguise, rather than asserting undeniable truths, the axioms are fixing the class of models that would satisfy them, therefore fixing the meaning[6] of the basic terms in particular applications. Thus, the relevant properties in the environment can be (functionally) demarcated without themselves being explicitly described. This model of an implicit definition for a mental representation allows for the dynamism, as well as for a progressive differentiation, construction, and precision of representations.

The aim of this thesis is to work out a deeper understanding of mental representation as implicit definition. The Interactivist model has considerable consistency and explanatory power, and yet it is not intuitively obvious. For instance, the analogy from definitions to representations requires explication. How adding new axioms in a formal system map onto additional differentiating criteria for dynamic representations, making them more precise, requires elucidation. Furthermore, this alternative understanding of representations brings up tensions between concepts and

---

dynamic. This is akin to the way, in linguistics, "presuppositions" are taken for granted without being present in the explicit propositional content of utterances.

[6] Here, 'meaning' is used in a particular way, the particularities of which is part of the subject of the overall thesis. What is delimited and determined here by the axioms are "relational meanings" whereby a model is constituted. As a standard definition of 'model' in mathematical logic, a model (of a theory T) is a possible realization of T that makes all the valid sentences (axioms) of T satisfied.(Suppes, 2002; Tarski, Mostowski, & Robinson, 1953) There may be different (and in some sense incompatible) realizations (models) that make a set of axioms satisfied (true). As a result, these different models might involve differing 'referents' for the basic terms. What is kept invariant between the different models is the (abstracted) relations between the basic terms. The relations are what are delimited by the axioms. Thus, a set of axioms may be realized by different geometries, the geometries would be the models, a basic term defined within the axiomatic system is foremost fixed with reference to its relational properties and only later by the referents that apply in each geometry.

properties (Carnap, 1988; Putnam, 1980), which leads us to functional role semantics (Brandom, 2001; Sellars, 1953a) and different understandings of functionalism (Bitbol, 2007, 2012; Block, 1980). Another point of explication is the difference between implicit and explicit representations, in terms of form and function, linking it to discussions on embodied cognition, surrogate reasoning and analogical representations such as maps.

The received view of representations has not only limited our understanding of representing, knowledge, and cognition. It has also limited the scope of questions and research subjects that are taken to be salient to our understanding of cognition. For instance, processes of learning and discovery have been deemed independent of knowledge and justification. The problem of origins and construction of content has not been taken seriously (Piaget, 1971). The possibilities and the limits of learning and discovery have not really been posed. Problem solving has been treated as a purely technical, mechanistic manner, and its ubiquity in learning and creativity is underestimated. Terminology and research areas that should be clustered and integrated, such as context effects, framing, premise selection; similarity, salience, and relevance realization; frame problems and notions of systematicity have been rendered apart and disunited. The process of interpretation and interpreting is not studied as part of cognitive science. Modal aspects of content that link creativity and productivity to systematicity, similarity and possibility are lacking. Functionalism likewise, is defined in a mechanical manner, normative aspects reduced to levels of explicit symbol conversions.

As suggested, the received view of representation is particularly vulnerable when we consider the origin and change of representations. This thesis takes the point of view

that the semantic, truth bearing aspects of representation and problems of justification cannot be solved independently from their context of emergence because representing is by nature dynamic and constructive (and reconstructive). Change of mental representations already implies cognitive construction, but our prevalent theories of cognition do not have an adequate account of the possibility of the construction of representations.

As a result, final aim of this thesis is to link and reinterpret these interrelated discovery phenomena, with the Interactivist account of mental representation in terms of implicit definitions at its center. The logic and nature of dynamic mental representations will be considered with regards to three phenomena discussed in philosophy of mind and cognitive science: nature and logic of representational content and representational change; skill development and dynamic learning in experience; function of homomorphic external representations (material symbols) and surrogate reasoning. With this, an improved understanding of learning from experience is hoped to be achieved.

Here is a brief overview of the following chapters. The current chapter was an introduction to the problem. In the next chapter (two), the general terms of the solution approach are laid down. Chapter three is one long argument running on for four subsections. It is laying down the pragmatist position, against the background of analytic philosophy. Chapter four deals with an illustrative example to demonstrate intricacies and implications in application. Chapter five is meant to point towards application and relevance of the approach toward further areas.

CHAPTER 2

IMPLICIT DEFINITION AS MODEL OF DYNAMIC MENTAL REPRESENTATION

The received view of mental representations is modeled after paradigmatic, external representations like maps, sentences, codes. Analogies and analogical models like images are also specimens of this notion of a typical representation. A common ontology among all these different forms of representations is presupposed. This involves content "carried" by strict forms (called vehicles), such as symbols or syntax, that can only be interpreted as part of a representational system of vehicles. For instance, a typical material symbol is an interpretable code that is assigned meaningful content, that can be both encoded (translated from content to code) or decoded (translated from code to content) because of a mapping between the contents and the vehicles.

Bickhard calls these kinds of paradigmatic representations "encodings", and the view that the nature of mental representations must be ontological encodings, "encodingism" (Bickhard & Terveen, 1995). Encodingism, more specifically, is the research program of starting with the premise of mental representations as biologically realized vehicles in the brain that stand in for semantic content and trying to cash out how this would (should) play out towards a coherent theory of cognitive phenomena. Therefore, in the literature, the quest for understanding mental representations in terms of their intentional and normative character as well as their physical material realization has been synonymous with "encodingism". The difference between the latter (the quest) and the former (program) is that the quest (which is some sort of naturalizing epistemology) does not necessitate the starting premise of the program, that "the physical vehicles carry representational content" (Shea, 2018). This premise of vehicles

that stand in for content is the landmark feature of encodingism: physical embodiment is encoding for this program. Vehicles are simultaneously type and token identifiable entities, and their structural properties do the work of standing in for semantic content, by means of converting[7] the structural properties into semantic properties.[8]

This premise is taken as a truism based on the very concept of a representation.[9] Taking it as a truism is what is disputed here. In other words, this premise is not necessary at all, for a quest to naturalize epistemology, or to account for the normative,

---

[7] Like a compiler in a computer program. However, a compiler, strictly speaking, is another translator "program", a code that converts high-level language program into machine language". Thus, the compiler deductively constructs an instruction code, but even in computing, simply compiling (converting/translating) is not processing. Even in the computer program case, a compiler program is written by a programmer and preinstalled, and its triggering is part of another program prewritten or manually triggered by the programmer. Any actions thus, are converted into machine instructions (decoded), to be "executed" (processed) by the CPU (Central processing unit or microprocessor). The CPU instructions, on the other hand, are executed by the arithmetic/logic unit (ALU) within the CPU and the order of the operations is controlled by the control circuit. The actual operations performed by the ALU are four arithmetic operations (simple addition, subtraction, multiplication, division), and logical decision-making comparisons (greater than, less than and equal to) and logic operations (such as OR and AND) that result in new (calculated) values being written in the necessary slots. The sequence and order of the operations is also controlled by means of the FLOW instructions, that tell the program to bypass some instructions, start from (go to) slots in the registry to execute. Strictly speaking, the FLOW instructions allow the programmer to intervene with the order of the code, that would otherwise run sequentially, to execute if – then conditions within the lines of instructions. Computer programming (computer algorithms) is about expressing possible human operations on a computer purely in terms of arithmetic and logical instructions, control flow among the operations and unique identifiers specified within the frame of reference of the computing space.
It takes a human programmer (engineer) to construct a computer as a usable deductive device, but programmer's construction is not a deductive operation.

[8] One may still argue that encodingism only applies to computational accounts of cognition, which is somewhat outdated. The core critique of encodingism also applies to many mainstream accounts that have notions of some sort of direct correspondence or isomorphism between an environment and representations in the brain (connectionism), as well as any notions of environmental "data" or "input" that is directly processed or "predicted by a brain" (i.e. predictive processing or Bayesian approaches) (Clark, 2013; Jones & Love, 2011); because such notions involve environmental properties somewhat "encoded" as biological stand-ins.
The aim in this section is not to criticize accounts, but to explicate the logical foundation of such accounts and the interactivist alternative.

[9] As well as the very concept of substance and efficient cause, as part of a substance ontology, as opposed to a process ontology. This connection will be taken up in the section on the interventionist account of causality.

epistemic as well as physical/biological and logical features of the agentive phenomena of representing. Indeed, according to interactivism (Bickhard, 2009), a process model of representation can achieve these goals more adequately. In this model, an agent's representations are an emergent property of the agent's anticipated internal flow. There is no homomorphism or direct correspondence. The actual "input signals" received from the environment are not representations, the data received are not epistemic inputs, as for instance in the predictive processing account where the brain "predicts the input". If one is talking about any sort of direct predicting of environmental input, such as predicting of data, predicting properties, predicting information in the environment per se, one is back to the encodingist paradigm.

There is a prediction (anticipation), but it is not directly encoding the environmental properties. Prediction is interactive flow of interaction indicating its next possible interactions. This kind of indicating as prediction is implicit (implicitly, rather than explicitly defined) in an organism's systematic capacity (and the related physiological and mental organizations that allow) to interact procedurally with an environment, which requires indicating possibilities of further action as well as selecting among possible actions. This indicating and selecting action possibilities is also implicit. The dynamic indications are similar to Gibsonian affordances,[10] where the environment directly offers the organism animal – environment interaction relevant possibilities of action (Gibson, 1979/2015, p. 119). A fly provides a possibility of tongue flicking and eating, for instance. Stairs provide a possibility of climbing for a toddler and a

[10] Strictly speaking dynamic indications are like *indications of* Gibsonian affordances, where a possibility of action is only indicated and is not explicitly represented.

possibility of leaping for a flea. This is also a very minimal understanding of representation that applies to many organisms. Possibility of control of interaction by the organism is already sufficient to capture foundational properties of representation by means of implicitly indicating possibilities of next steps in a dynamic, (environment) interactive flow.

But how so? Implicit differentiating is at the heart of this understanding of representation, but it is not so obvious what it means to differentiate implicitly, as opposed to explicitly. Indeed, the philosophical mysteries surrounding Gibsonian affordances in terms of their achieving the promised direct perception or "pick-up" are relevant here.

The question is how a property in the environment is directly picked up by an interacting organism. As Gibson points out, the affordances picked up are organism relevant, stairs as climbable by a toddler, but they are jumpable for a flea, etc. Why are affordances organism relevant? Because they depend on the levels of functional and structural organizations of the organism. It is not incorrect to say that being "jumpable" is an indication for the flea, however it is unnecessary (and indeed is contradictory) to postulate that being jumpable is represented explicitly as such. It is more correct to say that the indication is derivable from the possibility of jumping for the flea, which in turn is derivable from the capability of jumping for the flea, which is realized by the procedural flow of acting that the flea can realize, which can be *functionally* realized by means of an indexical control structure, which is further *functionally* realized over the physiological – chemical and structural organization of the flea. There are levels of functional realization, over other levels of physical and chemical realization.

Implicit defining makes functional realization possible, because a connected web of indices[11] together may function as an indication, so the function of indicating emerges. Indicating jump ability is implicit, implicitly determined in the capabilities that indicate possibilities for the agent. There are nested levels of functional emergence. Very simply, a procedural flow may be emergent from (may be realized by) a web of indices, and the flow of transitions between them. From such a functional organization emerge possibilities of interaction. When an index is pointed to, that index may indicate other webs of indices. More realistically, an encapsulated web of indices as a whole makes up an index for a "higher"[12] level. The organisation of feedback loops of indices and nested indices within them make up a capacity and possibility to interact.

Returning to the content – vehicle model, a legitimate question may be, whether in interactivism, we can still talk about "some sort of" vehicles, that "stand-in for" content. The objection would go like this: surely, there must be some sort of physical realization of a capacity to interact, and that relates to the content. So, why can't we say that the physical instantiations of the capacities, are vehicles of the content (for the interactivist model, as well)? When someone gains a new skill, their physical and functional organization changes, thus the way they interact changes, as a result the content of their representations changes. I will argue that the conclusion does not follow from the premise:

---

[11] An index is a pointer, like a signpost, that functions to point to another index. An index can be realized mechanically.

[12] The terminology of upper and higher levels assumes a foundational and an emergent level, as well as efficient causality. This vocabulary will ultimately be challenged in chapter 3 in favor of a more perspectivalist account.

Premise: When someone gains a new skill, their physical and functional organization changes, thus the way they interact changes, as a result the content of their representations changes.

Conclusion: The vehicle – content model applies to interactivism as well, whereby the physical instantiations of agentive capacities to interact are vehicles, and contents are realized over them.

We said that the signature of the received view is that it is committed to the vehicle - content model. Being committed to the vehicle – content model of representations is equivalent to being “encodingist”. So, if it were to be shown that interactivism also commits to the vehicle – content model, it would also be encodingist.

The vehicle – content model is more often termed the vehicle - content "distinction", but it also posits an identity between the (physical, physiological, chemical, structural, syntactic) vehicles and the contents associated with them. This is the core insight and solution of the theory to the problem of representation. The paradigmatic example is an individual symbol, which, my means of its structural features, acts as a vehicle with a causal role, while the same entity supposedly also has semantic interpretations. These accounts typically attempt to account for changes of representations and learning by means of changes to the vehicles. So that when the underlying vehicles are manipulated and changed, the content symbolized by them also changes. We must notice that the model supposes a direct and strict identity between them, because the individuation conditions of vehicles are tied to the individuation conditions of the content.

According to my understanding of the interactivist model, there is no ontological, structural identity relation between physical -causal vehicles, and content. If we push the

point to apply the distinction to the interactivist model, and take vehicles to be the physical, physiological, chemical, functional organizations that endow an agent with a capacity to interact in a controlled way with the environment, the vehicle-causally efficacious features of an agent by themselves are not representative. There is no representation, no content strictly speaking, in that picture. It concerns what is, what is actual in the agentive organization, which by itself could also apply to a thermometer.[13] The agentive organization is necessary for representation (the content) to emerge, but it is not sufficient by itself. In order for it to be representative, to have the semantic content bit, we need an ontologically different emergence, which is anticipation in the interactive model. Anticipation concerns possibilities, it involves recognition of possibilities by the agent. The possibilities (that make up the content) are a separate ontological emergence from the physical instantiation that enable the functional instantiation (the capacity). Notice that the word capacity connotes both possibility as well as actuality, both the content and the vehicle, and is therefore ambiguous. The main point is that, at the level of the actual organization, be it actual physical or actual functional organization, there is no representation. It is not sufficient for content. Vehicles do concern the actual level of instantiation, and that is not sufficient and therefore is not identified with content. The so-called vehicle – content distinction, strictly speaking, does not apply to the interactivist model. So strictly speaking, there is no vehicle ontologically and intrinsically identified with content.

---

[13] Or a carburetor, for that matter. The point is not merely being functional. One can say what makes a carburetor is to have a certain functional role, that it is a "being which mixes gas and air and sends the mixture to something else", which can be realized differently in different makes and types of cars (Block, 1980). Being functionally realized, having a functional role, mere functionalism is not sufficient for representation.

The received view is so intuitive, because there are ways to express it so that, it is merely an attribution of what is observed. If the received view is not claiming that the vehicle – content distinction explains what representations are, in other words, purely as a way of speaking, the received view is not strictly wrong. It is just saying that there are actual physical aspects of representations, which there are, and there are semantic, contentful aspects of representation, which there are. This gloss does not by itself give a theory of reference, nor a mechanism to realize that theory, unless one further attribute an identification between them. It could merely be a functional attribution. Since a lot of discussion goes back and forth between merely non-committal and trivially true ways of talking, and being committed, positing an encodingist identification between the vehicles and content, a standing for relationship, it is very difficult to argue against the received view. A similar point is made by Fodor, when he is criticizing what he sees as sense data verificationism, that representations are made up of simpler sense data. Incidentally, the view he ascribes to his opponent here is also encodingist, and he is making the point that encodingism is not so easy:

> Whereas, because the rest of us are not verificationists, we can live with the fact that 'chair' refers to chairs; we don't have to go around supposing that 'chair' refers to bundles of sense data. It is, of course, not very interesting to say that 'chair' refers to chairs, since we have no *theory* of reference and we have no *mechanism* to realize the theory. A fortiori, we don't know how to build a robot which can *use* 'chair' to refer to chairs. But, though "chair' refers to chairs' isn't interesting, we don't mind that so much since reference isn't what we're working on. We're working on logical syntax (psychologized). Moreover, "chair' refers to chairs' has one striking advantage over 'chairs are made of sense data'; it's not interesting, but at least it's *true*. (Fodor, 1978, p. 246)

It is true because it does not say anything. To the extent that it is trivial, one can admit it as true. Similarly, if the literature does not read into the vehicle - content attribution a theory and a mechanism of semantic content, one can reservedly admit it as true, almost

as a figure of speech. However, as a proper explanatory theory, it is problematic and inconsistent.

We said that the premise is correct to say that when someone gains a new skill, their physical and functional organization changes, thus the way they interact changes, as a result the content of their representations changes, but it does not end up with an identification between the vehicle and the content. The relation is that of emergence. We can get away from the strict identification of content with the vehicle because the identification stems from an explicit definition as a determinate stand-in, which models the logical form of an encoding. On the other hand, an implicit definition where a set of axioms implicitly fix the meaning of the axiom's terminology such as "line" or "between" in terms of the context (frame of reference) and the conditions under which they would be made true, break the identification of the conditions with the meaning, allowing ontological emergence. Let us highlight the contrast between explicit definition, as opposed to an implicit definition further. The preceding part will proceed from the contrast of implicit and explicit in formal systems towards the model's application to dynamical representations of interacting organisms.

2.1 Explicit definition in formal systems

An explicit definition is a definition given by an equivalence relation. Explicit definitions are prototypical form of definitions: they are the form of stipulative formulae and stipulative terminology, they are the ideal for definitions in a dictionary to converge towards.

Explicit definitions are said to satisfy the conditions of eliminability and non-creativity. Non-creativity (sometimes also called conservativeness) refers to the fact that

"new things should not be provable by way of the use of definitions" (Gupta, 2011, p. 88). A prototypic example is an abbreviation, which is a perfect stand-in. A non-creative definition is also called "'normal' or "proper' or 'purely verbal" (Popper, 1964). Non-creativity (conservativeness) is the condition that any surplus of meaning between a definiendum and a definiens is avoided when defining an expression or when introducing a word or symbol into a logical system by providing a statement of its meaning, whereby an equivalence is established.

Eliminability refers to the criteria that any sentence containing the symbol to be defined must have an equivalent sentence that doesn't contain the symbol. A definition of d should contain neither the expression defined in d, nor any expression defined with its help (Popper, 1964).

Explicit definitions are said to satisfy *Realistic eliminability*[14] the same meaning is expressed on both sides of the equivalence, for instance:

$y = f(x)$

A bachelor is an unmarried man.

"BBs" stand for "black balls".

Non-creativity (conservativity) condition is linked to the criterion of necessity so that nothing unnecessary to the definiendum (that which is defined) is introduced by means of definiens (the definition). Eliminability is linked to the criterion of sufficiency such

[14] Implicit definitions are said to satisfy anti-realistic eliminability: a definition is a set of rules, meaning is the conditions under which an expression can be asserted or proved. (Poggiolesi, 2012) Compare this to the difference between explicit and implicit *equations*: an *implicit equation*, is an equation which relates the variables involved. For instance, the implicit equation "$x^2 + y^2 = 4$" gives a relationship between x and y, even though it does not specify y explicitly in the form y=f(x). As in this example, it may not be possible to convert an implicit equation into an equation of the form y=f(x) for some function f(x), as the x values in the range $-2 < x < 2$ each correspond to two values for y.

that the definiens (the definition) is sufficient to eliminate in place of the definiendum. So, the criteria give us the necessary and sufficient conditions. In other words, explicit definitions express *an equivalence* of meaning between the definiendum and the definiens, one is fully eliminable in terms of the other, defined fully (sufficiently) and conservatively.

Consequently, explicit definitions are always (by definition) *stand-ins* for other definiens, eliminable in terms of the other: x = y = z = w ... and so on. Explicit definitions are denotational, meaning direct. This is why they are called "explicit".

But where is their origin? How do these equivalences emerge, and where do these equivalences come from in reality? The origin of the definitional equivalences cannot be due to the definitional relationship, due to non-creativity, their origin is not intrinsic to them. Their origin must be provided or introduced as stipulated by an interpreter within a system of symbols, language, or convention. But what does the interpreter do in interpreting an explicit definition for a definiendum, cognitively? On pain of regress, interpreters must translate the definiendum into directly consumable, intrinsically meaningful representations which are not merely stand-ins themselves.

The logic of these direct representations is modeled after implicit definitions, definitions that demarcate without an equivalence.

2.2 Implicit definition in formal systems

Implicit definition is an alternative to denotational explicit definition. Implicit definition is most well-known with Hilbert's axiomatization of Euclidian geometry. The axioms are statements that collectively specify the minimal conditions for the existence of mathematical or scientific systems. Some well-known examples of implicit (structural)

definitions are: the axioms of Peano arithmetic serve as a definition of the primitive arithmetical terms ‘natural number’, ‘zero’, and ‘successor’; ZFC defines the membership predicate ‘∈’ in the language of set theory; Hilbert’s axiom system of Euclidean geometry serves as a definition of the primitive geometrical terms ‘point’, ‘line’, ‘plane’, as well as of the relation symbols ‘incidence’, ‘betweenness’, ‘congruence’, etc.

Hilbert divided the extra logical terms used in his axiomatic system into 1) primitive (basic) terms not defined in the axiomatized system and 2) defined terms, which were explicitly defined by means of the primitives. Thus, in his axiomatization of Euclidean geometry, the terms 'point', (straight) 'line', 'being between', are among the primitives, while 'line segment', 'angle', 'triangle', 'length' is among the defined terms (Hempel, 1951). No further explicit definition is provided for these “undefined” terms. Instead, they are said to “implicitly define” their satisfiers as the classes of things which make the axioms true, in a particular interpretation. The nature of these primitive objects and the relations between them are arbitrary if they satisfy the axioms.

Thus, a common set of axioms could result in different physical interpretations of line segment, depending on the additional axioms that determined whether the plane in question was a Euclidian plane, the surface of a sphere or the surface of a saddle. For instance, only denying the parallel postulate (axiom) from Hilbert’s original set of 20 axioms allowed other consistent geometries to be constructed:

> Let there be given a straight line a and a point *A* not on that straight line.
>
> *The Parallel axiom:* Then there exists no more than one straight line passing through *A* not intersecting a and lying in the plane defined by a and *A*.

Alternative axiom 1: Then there exists no more than one straight line passing through *A* not intersecting a.

Alternative axiom 2: There exists more than one straight line passing through *A* not intersecting a.

Alternative axiom 3: There exists no straight line passing through *A* not intersecting a.

Notice that nothing is explicitly stated with regards to the actual description of the line segment. Nevertheless, specifying the domain of interpretation specifies and determines the structural properties of the line segment in each domain. Figure 1 illustrates this point through the domains of application of alternative geometries. When curvature is not specified, an overlapping set of axioms may apply to different domains, ending up with different realizations of straight line.

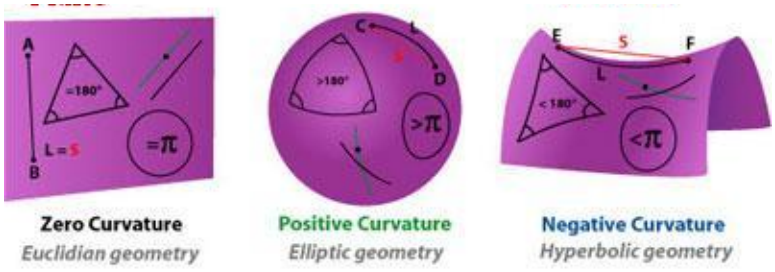

Figure 1. Domains of alternative geometries

Hilbert (and Schlick and Peano) were concerned with a particular kind of rigor, in terms of not stating more than that which can be rigorously defined. They refrained from explicitly including "intuitive" or "self-evident" properties of the objects as explicit definition unless it was tied to the axioms, which led to the discovery of alternative geometries. "Schlick had championed implicit definitions that laid down "the mutual relations of the primitive concepts [of a theory] as expressed in the axioms" (Schlick,

1974, p. 36) as fixing the meaning of scientific concepts without reliance on intuition, in terms of their formal relational properties alone." (Uebel, 2016)

On the other hand, Frege had different criteria for rigor. He objected that implicit definitions are not definitions; they merely define equivalence classes. Frege was interested in deductive proof: "The need to exclude with certainty any tacit presupposition in the foundations of mathematics" (Currie, 1982).

Like many debates in the history of philosophy, the Frege – Hilbert debate was concerned with, what is more important. Hilbert was not looking for simple and certain, unwavering rules that give truth, instead, he was looking into the origins and the deeper connections, the deeper relations between the axioms, a safer, more fail free way to define:

> He told Frege that the purpose of Hilbert [1899] is to explore logical relations among the principles of geometry, to see why the "parallel axiom is not a consequence of the other axioms" and how the fact that the sum of the angles of a triangle is two right angles is connected with the parallel axiom. (Frege - Hilbert Letters, translated in Shapiro, 1997, p. 162)

Hilbert's aim was to be free of any intuitive notions that might be mistaken. He wrote "I do not want to assume anything as known in advance." (Frege - Hilbert Letters, translated in Shapiro, 1997, p. 162). The definition is not made explicit, it is left implicit, up to the structurally, isomorphically necessary level whereby schematic concepts (terms, variables) and their relations to each other are defined. They differed on the criterion of truth. Hilbert "repeated the role of what is now called "implicit definition," noting that it is impossible to give a definition of "point" in a few lines, because "only the whole structure of axioms yields a complete definition."

Frege worried about the ideal elements introduced into geometry (by means of implicit definitions), and wanted to reach at truth, then consistency would follow. Hilbert did not rely on the truth of individual axioms based on the definitions and their intuitive truth based on spatial intuition (Euclid's geometry) and wanted to rely on logical consequence only. For that purpose, "if the arbitrarily given axioms do not contradict each other with all their consequences, then they are true and the things defined by them exist" (Frege - Hilbert Letters, translated in Shapiro, 1997, p. 162). This way of expressing a link between formal consistency and existence was problematic, as criticized by Frege:

> Hilbert said that if we establish the consistency of his axiomatization, we thereby establish the existence of points, lines, and planes. If we establish the consistency of the theology axioms, do we thereby establish the existence of a God? (Frege - Hilbert Letters, translated in Shapiro, 1997, p. 163)

Hilbert was referring to "abstract structures", but Frege did not think in terms of "truth in a model." For Frege, the quantifiers of mathematics range over *everything*, and a concept is a function that takes all objects as arguments. Thus, "my pocket watch is a point" must have a truth-value, and our theory must determine this truth-value. According to Frege's position, "the only way to establish consistency is to give a model: "to point to an object that has all those properties, to give a case where all those requirements are satisfied.", to avoid any issues with exemplification.

Here Hilbert is anticipating the structuralist project:

> [A] concept can be fixed logically only by its relations to other concepts. These relations, formulated in certain statements I call axioms, thus arriving at the view that axioms . . . are the definitions of the concepts. I did not think up this view because I had nothing better to do, but I found myself forced into it by the requirements of strictness in logical inference and in the logical construction of a theory. I have become convinced that the more subtle parts of mathematics . . . can be treated with certainty only in this way; otherwise one is only going around

> in a circle" (Frege – Hilbert Correspondence, p. 51). … An implicit definition characterizes a type of structure—a single structure if the axiomatization is satisfiable and categorical. (Shapiro, 1997, p.164)

Hilbert was defining freestanding structures, while Frege wanted to give true definitions to terms we use in our theories. A lot of criticism, in accord with Frege, was with regards to the interpretative status of the implicit definitions, which we will talk about in the next section.

The first important take away from this discussion, in terms of modelling mental representations, is that "equivalence classes", or "models" functionally defined is exactly what we need for mental representation. Explicit definition model of mental representation is inconsistent and cannot account for many properties of mental representations. If mental representations are to be successful differentiators of action relevant properties, as well as originate in the brain, they must allow demarcation without description; and implicit definitions seem to be able to meet this feat.

The second important take away pertains to the conversion from implicit to explicit definitions. There are ways to make implicit definitions explicit or denotational, by means of adding further premises or specifying domain of interpretation (domain of reference), that one may constrain the resulting equivalence classes without explicitly describing what is represented, as in the line segment example. It is advantageous (more rigorous) to be able to keep definitions vague up to the point where it becomes necessary to specify them. This also allows progressive differentiation and getting more precise. This property of implicit definitions is also desirable for a model of mental representations which are supposed to get more precise with increased expertise.

As a result, implicit definition is practically stronger and more conservative (and less error prone) because the organism starts with just the minimum constraints, the functions and the roles that the object must fulfill, to satisfy its internal process requirement more natural ground for representations, whose aim it is to avoid error. We can say that implicit definition as a logical model is able to satisfy the dynamic properties of mental representations, better than explicit definition would.

2.2.1 Types of implicit definition

What is called implicit definition or 'definitions by axioms' in the literature are actually two different kinds: implicit definitions in the strict sense in the style of Gergonne, and structural definitions in the style of Hilbert (Giovannini & Schiemer, 2021). Shapiro defines strict implicit definitions as follows:

> In strict implicit definitions, all but one of the terms in a language has fixed meaning…when there is exactly one way to assign a denotation to c to make every member of set of sentences S true. So, for example, "c is a perfect number and $c < 10$" is an implicit definition of 6 in arithmetic. (Shapiro, 1997)

What we are interested in are structural definitions. They do not uniquely define. They define partially. Shapiro calls them also "functional" definitions whereby "an implicit definition is a *simultaneous* characterization of a number of items in terms of their relations *to each other*" (Shapiro, 1997, p.130).

Structural definitions are sometimes attributed to be contextual or "usage" definitions[15] as Russell named them (Gabriel, 1978). This attribution is usually

[15] "Usage definitions define an expression by giving meaning-equivalent reformulations in which the expression no longer occurs for statements or statement schemes in which the expression occurs. The meaning of the expression to be defined is thus determined in the context of other expressions that have

unfavorable, to argue the point that structural definitions are not really “definitions”, because they cannot be solved unambiguously or deductively: “an axiom system in Hilbert's sense may (should) admit several solutions, i.e. models” (Gabriel, 1978). This is contrasted to the way Gergonne’s “system of several such statements forms an implicit definition. In analogy to systems of equations with several unknowns, Gergonne assumes that the unknown meanings can be determined with the help of the known meanings. Accordingly, he provides that the number of unknown meanings corresponds to the number of statements.” (Gabriel, 1978) Strict implicit definitions are conservative, they are not creative. They are eliminable through deduction.

Gabriel (1978) contests the claim that “the definitions by axioms appearing in Hilbert's Foundations define the basic predicators like 'point', 'line', 'plane', 'between', etc.”, he attributes this “confusion” to structural (contextual) definitions being taken for Gergonne type strict implicit definitions:

> The so-called definitions by axioms do not define the basic predicators of the axiom system since they do not fix four meanings. Hilbert's axioms are not statements, but forms of statements. In these, the basic predicators are not predicators with specific meanings, but variables for single-place predicators (e.g. 'point') and multi-place predicators (e.g. 'between'). (Incidentally, these variables are not free variables, but merely identify argument places.) What is determined by axiom systems in Hilbert's sense are therefore (in analogy to Frege's way of expression) not the first-order basic predicators, as Hilbert initially takes it seemed, but (multi-digit) predicators of the second order. An axiom system with n (alleged) first-level basic prs determines an n-ary second-level predicator - a structure, as we say today. (Gabriel, 1978)

---

already been defined. Therefore these definitions are also called ' context definitions '. The best-known example of such a definition is Russell's definition of the labeling operator in the Principia mathematica.” (Gabriel, 1978)

> Accordingly, while the definitions by axioms only specify formal structures (multiple predicators of the 2nd level), the implicit definitions should in fact define predicators of the 1st level. (Gabriel, 1978)

> Through the implicit definitions in the sense of Gergonne - at least according to the claim - the individual unknown expressions should receive a certain meaning, i.e., the implicit definitions should in principle be resolvable into explicit ones. However, this does not apply to the usage definitions. They are not intended to determine the meaning of individual expressions, but of the entire statement in which the expression appears in the context of others. (Gabriel, 1978)

The importance of the position espoused by Gabriel, along with Russell, and Frege, lies not in the formal features of implicit definitions, and what can be deduced from them, but in the ontological status they attribute to implicit definitions. To them, implicit definition is simply a level of abstraction, specified derivatively from the level of predicators, which is the real ontological level. As a result, the implicit level, as a level of abstraction, is not an ontological level, because it is uncertain, which they take to be undefined.

On the other hand, according to the structuralist view of Shapiro (1997), implicit definitions describe ante rem structures: metaphysical universals, as opposed to physical or logical universals. It means that structural relations are not merely a level of abstraction from the level of objects, they have a metaphysical reality of their own. As a result, the ontological status of relational structures, also concern the ontological status of objects, relative to them. In philosophy of mathematics, structuralism is the view that relational structures are prior to the mathematical objects. Mathematical objects, as "singular terms [are used] to denote the places of the structure" (Shapiro, 1997, p. 130)

Both camps agree that implicit definitions specify relational structures, what is at issue is the ontological status and independent reality of these structures. As a result, the

structuralist position which gives ontological reality to structures perhaps specified by implicit definitions is more congenial to the ontological reality of representations as functionally abstract structures specified by implicit definitions.

The certainty and resolvability that is lacking in Hilbert-type, structural implicit definitions enable a gain in expressive power. It is true that (structural) implicit definitions are not always resolvable into explicit definitions. This, however, is a power of expression[16] when it comes to being able to represent, stand-in for, unbounded and intractable features of ordinary human experience, that cannot be captured by explicit definitions. This is one of the reasons, attempting to represent contextual features ends up in frame problems. Since human cognizers do not have such a problem, their mental representations cannot be modelled after explicit defining.

2.2.2 Implicit definition and unconditional assertability

A related discussion in the history of philosophy concerns implicit definition and unconditional assertability. For instance, in "Implicit Definition Sustained" Quine rejects taking implicit definition as a source of analyticity; that is taken to be analytically true (unconditionally assertable) with respect to implicit definitions (W. V. Quine, 1964). In the same vein, Christopoulou (2013) raises the problem of the synthetic content of implicit definitions, if it is true that implicit definitions are "true by stipulation". Christopoulou raises the problem with regards to three kinds of applications of implicit

[16] This aspect of knowledge representing is problematized in computing literature: for representational formalisms, there is a trade-off between expressiveness (expressive power) of a representational language and its computational tractability, and indiscernibility. This is posed as a limit on automated artificial reasoning (Davis, Shrobe, & Szolovits, 1993; Levesque & Brachman, 1987; Orłowska & Pawlak, 1984).

definitions: "postulates of scientific terms, mathematical systems of axioms and abstraction principles." She argues that none of these applications are purely stipulative (a priori), they are either empirically, semantically, or metaphysically factual, and thus accounting for their facticity along with their stipulative aspect is required. Sellars puts the problem as the objection against the definitional character of implicit definitions: “"Implicit definition," it is pointed out, is a purely syntactical affair, and to expect it to give rise to extralinguistic meaning is as sensible as expecting a number of people to lift each other by their bootstraps.” (Sellars, 1953b, p.125) Sellars responds that this objection applies equally to explicit definitions, and points out an essential feature of language; that it is rule based:

> But its force as an argument against the definitional character of implicit definition is somewhat less keenly felt when one realizes that exactly the same limitation holds of explicit definition. Both explicit and implicit definition are matters of syntax. The difference is that whereas in the case of explicit definition the definiendum and the definientia are distinct, and the 'giving extra- linguistic meanings'-however this is done-to the definientia fixes the extra- linguistic meaning of the definiendum; in the case of implicit definition the extra-linguistic meaning must be 'given' to all the predicates 'simultaneously', as they are all both definienda and definientia rolled into one. (Sellars, 1953b, p.126)

By “matters of syntax”, Sellars means inferential rules. Both explicit and implicit definitions are defeasible, even though explicit definitions are more restricted. Defeasibility and being interpretable is not relevant to their utility as ‘definitions’.

> In any event, to the fact that the syntactical structure of a chain of explicit definitions limits the number of alternative real meanings which can be possessed by the predicates in the chain, corresponds the fact that the number of possible 'interpretations' of a set of implicitly defined terms can frequently be narrowed by adding a new axiom to the original set. In neither case would the utility of the definition seem to depend on its admitting only one set of real meanings.

The rules, inferences, consequences of the axioms limit further inferences which can be defined from the terms, therefore defining it. In explicit definition, substitution (eliminability) is more salient. In implicit definition, different models can satisfy and therefore give truth, there is no direct substitution. Direct substitution is an achievement of more developed differentiation and discerned understanding: it requires knowing the conditions (contexts) of substitutability, in other words, knowing under which contexts unconditionally assertable. If truth is unconditional assertation (within a frame), explicit definition is more generalized (more likely to be unconditionally asserted) than implicit definition. Implicit definition separates meaning from unconditional assertability, takes up meaning as permissible inferences.

2.3 What would such dynamic representations be like?

According to interactivism, the dynamic representations that such a logical model would support would be dynamic indications defined implicitly, that function similarly to Gibsonian affordances. Three aspects of the representations will be explicated: 1) their dynamism, changeable nature 2) their semantic and epistemic content 3) their causal role and causal construction and sustainment.

2.3.1 Encodings are by nature static

We have seen that the received "encodingist" view of mental representations that logically model explicit definitions depict something like corresponding (homomorphic) projections of environmental properties in the representational system of the organism. Some sort of direct correspondence or homomorphism is indispensable for such accounts because this is how they achieve faithfulness to facts and realism. The

assumption that factual accuracy requires correspondence with the environment is the suppressed premise in these frameworks. Even in the most sophisticated versions, the homomorphic (or isomorphic) correspondence can (and should) be translated into an equivalence between the environmental properties and the organism's representations of these properties so that the agent can successfully interact with its environment based on these accurate representations. And the equivalence is logically modeled after explicit definitions.

The indispensable suppressed premise introduces problems for dynamism of mental representations. A procedure which creates corresponding representational elements in an agent's representational space must be some sort of "copy" procedure. If the aim of such a procedure is to create some sort of isomorphism, some sort of correspondence, some sort of faithfulness or covariance with the environment, that means the source and criteria of such a procedure is the environment from which different methods of copying[17] are at stake.

Normally, a "copy procedure" may involve a mechanical "construction", for instance a 3D printer printing a replica organ slowly during an extended period is described as "constructing an organ". Compare this to a computer programmer constructing a software code. The former is not construction in the metaphysical sense, where a new form, a new organization with new functional properties comes into being, while the latter is. In the mechanical construction case, the same form is constructed

[17] As an empiricist understanding of representation, this account is subject to the problems of empiricism such as "the myth of the given" (Sellars, 1956/1997). The difference here between a blind copy procedure and a metaphysically true constructive procedure concerns the difference between Humean and later positivistic understanding of efficient cause and an alternative Aristotelian agentive or contemporary interventionist cause (Bitbol, 2012). This link will be explicated in the section on functionalism and interventionist theory of causation.

from new materials, so the form is “copied”. The copy is faithful to the form of its source, but what warrants its correspondence to (the form of) the source is the blind copy procedure. The procedure would involve a regularity of sequence, event to event triggering and succession which must be invariably blind to be causally reliable. The output of the procedure is (and should be) fixed so long as the input is the same. It is an equivalence, an explicit function in the form of x= f(y). A copy is a “copy of”, therefore it is already fixed, static, and therefore actual and factual. For a copy, being reliable and being static go together.

According to Bickhard (1980, p. 8), another associated inconsistency with the copy concept, as it applies to information in particular, is the transmission model of communication. Not all (iso)morphisms entail transmission models, because there may not always be a transmission assumed. This is a separate issue from the one above. Here again a copy of the “message” in its formal or coded form as a signal is thought to be sent as encoded – transmitted as copied form – and received as decoded. In encoding models, “extracting information” from the environment may assume such a transmission model, where the transformation into agent relative information is simply assumed, without being explicated. Taking such a model as explanatory is inconsistent because the transformation of underlying knowledge into decoded signals (and vice versa) requires rules of transformation, it is precisely the rules that require explanation. Such accounts confuse the actual token of a physical signal with its interpreted meaning with regards to a system of units of differentiation, which is a type or an abstraction. Another way this gets confused is with regards to the difference between Shannon (mathematical) information and semantic information: the former is derivative from the latter, in the sense that any meaningful differentiation can only be realized as a causally reliable

physical interaction by means of systematic and abstract relations of control between systems. Exerting control, in the sense of influence or determining, allows the realization from physical tokens to functional control. The measurable unit of meaningful physical tokens is mathematical information, which is derivative from the system of measurement, as a system of control.

> [T]he concept of information is subsidiary to that of control in the sense that any signal or structure contains information relative to a particular system only insofar as it is capable of exerting control on that system. Mathematical information is a measure of the degree of potential control. (Bickhard, 1980, p. 8)

Measure theory talks about how one captures empirical conditions that reliably occur under well-defined conditions and map the initially qualitative comparisons into quantitative magnitudes: values within a well-defined range and unit system of values. The mapping is essentially a transformation (and not a transmission) from qualitative differentiations to quantitative differentiations, made possible by drawing out two homomorphic systems.

> Numerical measurement is the systematic assignment of numbers to objects in order to represent certain of their properties, specifically those properties, which we call 'magnitudes', that are capable of instantiation in different quantities. Such assignments are *possible*, because the magnitudes to be measured *satisfy certain well-defined empirical conditions, which are the possibility conditions* of a measurement practice for the magnitudes in question. (Matthews, 2010, p. 128) (my emphasis)

> These values are determined by certain fairly elementary empirical procedures. In the case of mass, for example, one can use a single-arm pan balance to establish either that two objects have the same mass or that one has a greater mass than the other. By successively concatenating objects of unit mass, one can construct a scale for arbitrary whole multiples of unit mass. This scale can (subject to certain practical limitations on weighing operations) be arbitrarily refined by using the pan balance to identify objects of mass 1/n (in effect by finding n objects of identical mass that together have unit mass) and then successively concatenating one or more of these objects both with themselves and with the objects of whole multiples of unit mass to construct a more fine-grained scale. (Matthews, 2010, p. 130)

When it comes to semantic vs mathematical information, a similar relation holds: the possibilities of control exerting relations (influence or determine) which determines the empirical consequences are converted into the amount of control with the mathematical information. There is a transformation involved to get the homomorphic mapping from the qualitative empirical interaction domain to the quantitative domain. As a result what deems the difference meaningful and significant (instead of mere "noise") is that it is "a difference that makes a difference" (Bateson, 1973/2000, p. 428), or more precisely "a difference that *can* make a difference" (Sloman, 2018). "[A] difference, then is an abstract matter" because it needs to be interpreted as difference (Bateson, 2015, p. 93).

There is nothing explanatorily dubious with a copy procedure which can be designed and used by a human agent. Problems come up when the same procedure is assumed to explain the way mental representations originate in perception and cognition. It is more of an abstraction, an explanatory gloss (Egan, 2014).[18] The homomorphic mapping is not the problem; assuming the homomorphic mapping already achieved, without any interpretation required, disregarding the conditions of the possibility of achieving the transformation into a homomorphic mapping from one domain of interaction to another domain of interpretation is the problem. For instance, for the output to get more precise, to become a more faithful copy of the source, the copy procedure must change. The blind procedure in its current stable regular form cannot provide more accurate, more precise outputs than it is already producing.

---

[18] "The mere existence of the causal relation (or of an appropriate homomorphism) would not itself justify the content ascription in the absence of the appropriate use" (Egan, 2014). The act of ascribing the content is giving the phenomena its intentional character and not the mechanical procedure described.

This case applies to any change of mental representations through learning and skill development. A "copy" procedure cannot result in gain in precision, as it is already fixed, actual and factual. Similar considerations would apply to any notions of "innate" static representations that directly represent environmental properties as encodings. The procedure that supposedly selects out these innate representations may be dynamic over time, but that mechanism is irrelevant to the ontogenic development of these innate representations. That would be a separate process and would also require true construction in the metaphysical sense (aka., emergence). Since innatist accounts do not really have an explanation for change of representations, they cannot account for the original emergence of these innate representations that were selected in for. Encodingist notions that are modelled after explicit definitions, be they empiricist or innatist theories, categorically cannot account for the change and dynamism of mental representations.[19]

2.3.2 Dynamic differentiations implicitly defined

When we take the nature of mental representations as implicitly defined dynamic differentiations[20] modelled after implicit definitions, accounting for the construction and change of these mental representations becomes natural. The properties in the environment that are relevant to the successful course of an interaction are not explicitly represented as criteria or conditions. Instead, properties are specified and determined *implicitly*, just like the shape of the line segment is determined based on the domain of

---

[19] For related reasons, misrepresentation (possibility of inaccuracy of representations) and agent detectable error (the agent detecting the inaccuracy of their own representations, the ground for the possibility of saying "I was wrong") is antithetical for static representations.

[20] Instead of encodings.

application, whether of a Euclidian plane, the surface of a sphere or the surface of a saddle.

Notice that the abstract entity of a line segment is only one of an unbounded number of entities that can be thus specified in the environment. Each possible entity that satisfies the stated axioms within that space (say possible types of triangles) does not have to be enumerated and designated, to be specified by the implicit definition. In fact, enumerating and designating would fall into representing explicitly. If set of entities that satisfy an implicit definition are to be represented more specifically, they can be further defined either implicitly or even denoted explicitly. To denote anything explicitly, what is already implicitly defined must be uniquely determined in the context of a frame of reference.[21]

As a result, what is defined by an implicit definition is not unique. The class of items that satisfy an implicit definition may be unlike each other in some ways: black pebbles as well as flies may satisfy the same representation for food for a frog; "[a]n indication for the frog, for example, that there is a tongue flicking and eating opportunity is unbounded with respect to the number of actual flies, or fly conditions (or tossed BBs), that would support it" (Bickhard, 2023).

Furthermore, the class of models that satisfy the implicit definition is unbounded. There is no a priori limit on the kinds of objects that would satisfy the axioms, apart from the minimum conditions that the axioms specify. This was the rigor

---

[21] In this model, the notion of a *particular* is a derivative and an achievement which can only be fulfilled based on the availability of criteria of individuation, criteria of identification and criteria of reidentification (Bickhard, 1998). Otherwise, everything is general up to pragmatically allowable scales: "…one may fail to appreciate that not only is the mapping between representing and represented structures non-unique up to admissible transformations of scale (as specified by the uniqueness theorem), but also that the choice of representing structure will itself be non-unique" (Matthews, 2010, p. 139).

of Hilbert's program. No intuitive, prototypical, or self-evident properties are predicated, independent of the axioms. There are no explicit descriptions of properties that would satisfy the axioms. That part is undefined until further axioms are added that specify them. The equivalence class that an implicit definition points to may be bigger (more general) or smaller (more specific), depending on the axioms. The argument here is that mental representations do not require unique reference or denotation. On the contrary, the fact that implicit definitions specify equivalence classes is a virtue to allow change and progressive differentiation of representations.

2.3.3 Interactive notion of representation

If the minimal condition of defining implicitly is sufficient for specifying equivalence classes, the idea can be generalized to differentiating implicitly as necessary aspects of interaction. A dynamic differentiation is a generalization of the concept of implicit definition to apply to the pragmatic context of agentive action flow. Interaction requires capacity to control interaction by the organism. Possibility of control of interaction by the organism at a minimum involves action indication and action selection, both context (environment) specific. This can be swimming or tumbling action for a bacterium toward increased sugar. Action indication and selection by an organism in an actionable environment is already sufficient to capture foundational properties of representation in terms of causal covariance, truth value and normativity, without having an encodingist "image representation" of environmental properties.

> [C]onsider a subsystem engaging in interaction with its environment: the internal course of that interaction will depend both on the organization of the subsystem and on the interactive properties of the environment being interacted with. That same subsystem interacting with a different environment might engage in quite different internal processes. (Bickhard, 1998, p. 186)

The outcome of the interaction yielded will be jointly determined by the organization of the functional system in the organism that is engaged in the interaction, and the properties of the environment being interacted with. In this sense, possible final states of interactions may suffice to differentiate between and therefore serve to classify types of environments (Bickhard, 1998). As a result the organization of the internal process flow of the organism implicitly "defines" the class of supportive environments without encoding the environment properties per se. Internal process flow is constructed and sustained based on past experiences. Internal process flow covaries and is coupled with the appropriate environment to the extent that the agent is fluent in that environment. The existing functional organization of the agent allows for action selection. This is the actual, factual, and causal link between the organism and the environment. It has two more aspects for representation to emerge.

The second aspect is the truth value of representations, which is tied to the concurrence between the associated internal outcome of an action flow and the actual outcome. A glitch or interruption in the flow provides a falsification. Smooth flow of action is a corroboration in the Popperian sense. What is falsified is the further indication of interaction, and conversely what is corroborated is the existent indications of action flow in the specific environment applicable. While the truth values of the representations are either confirmed as successful or falsified as failed, the deemed (implicitly indicated) environmental conditions holding in the environment are also either confirmed or falsified.

Error detection (and falsity), however, is not necessarily scrutable since the interactive flow is corroborated or falsified as whole. We say that classification of types

of environments is implicitly defined, based on possible final states of interactions, without ever being explicitly about the environment. Only the information is functionally available. Which properties in the environment must be true for the representations to hold true is not specified explicitly. As a result, the reason for the failure is not scrutable initially, but it can also be progressively functionally differentiated, based on further dynamic interaction. Furthermore, these truth conditions are unbounded. An unbounded number of possible environments may satisfy the conditions for successful action, and all are covered in this way.[22]

The success of the interaction flow is dependent on both the internal process flow and particular conditions being the case in the environment, because ideally the two should covary. When the process flow fails, it may be a failure of the organization of functional process flow (competence), or it may be a failure of indicating the appropriate

---

[22] One may pause the question, if all representations have truth value, how representations of environmental properties can have truth value (or can misrepresent), without having the form of a judgement or a proposition. Since the form of mental representing are not explicit encodings, implications of compositionality (of judgements into sub-elements), does not directly apply to representing in this model. Compositionality is not intrinsic to the form of truth value functions; it can be rendered derivative from it.

What may be said to apply is the model-theoretic interpretations of truth. Here all representing is "functional": involves truth value functions, always interpreted as satisfied or true. Explicating its nature would go into explicating truth in varieties of structuralism, which involves discussions about purely stipulative vs factual aspects of implicit definitions, as well as the link between implicit definitions of class models and implicit definitions of terms. For instance, if we take relativist structuralism: modeling reference on the notion of interpretation in model theory, where truth is interpreted relative to the model that is initially chosen, where the arbitrariness of the chosen model is unimportant as long as we keep consistent with the model. Because all appropriate models would be isomorphic with respect to the relevant conditions. Although there is a variable aspect of the models which end up with something akin to "referential ambiguity", the relational - configurational (isomorphic) aspects are kept consistent as long as we are consistent with the model used: "But why does such relativity of reference not cause problems, in particular with respect to truth? The answer is, of course: because all models of P A2 are isomorphic. Thus we will always agree on the truth value of a given statements in the language of arithmetic, no matter which model we have picked initially. In other words, while truth has been *defined* in a relative way, a non-relative notion of 'truth in arithmetic' is actually *implied*: truth in all models of PA2." (Reck & Price, 2000, p. 350)

process flow (a failure of applying the competence). This leads us to the third, the epistemic and contentful aspect of representations:

> Engaging in interactions is functionally anticipatory… the success or lack thereof, and the conditions upon which that might depend, are in the future of the interaction once initiated. As such, there is an implicit predication that "this" is one of those environments in which the initiated interaction will proceed as anticipated. That predication, therefore, might itself be true or false: the environment might or might not be among the supportive kinds. (Bickhard, 2009b, p. 569)

As a result, there is a presuppositional, therefore implicitly predicated, functional relationship between the activities of the organism and the environmental conditions. This constitutes minimal representation. Representation implicitly emerges in the indication and selection of interaction possibilities, as anticipations of the organism. For an action possibility to be indicated, particular dynamics of interaction should be available to the organism. However, the availability of dynamics is not sufficient for representation. The implicit presuppositions pertain to certain conditions holding in the subprocesses of the organism, or in the environment, it involves all necessary conditions for the interaction flow to proceed as anticipated. The issue of the semantic and normative aspect of representations arises because the anticipations with regards to implicit predications may not hold. Implicit predications not only are satisfied or dissatisfied, but they also serve to (implicitly) differentiate between suitable conditions being the case in the environment. Thus, anticipatory indications of interactive possibilities constitute *representations*. The content is provided by ‘‘forward looking’’, modal indications of potentiality. If, anticipatory indications are falsified, by a violation of the course of an interaction, and the interaction does not proceed as anticipated (the implicitly predicated follow-up interactions cannot pursue), or exceeds the "bounds of

what is functionally anticipated, then that anticipation is falsified – for the organism itself" (Bickhard, 2009, p. 570). The model applies not only to triggered interactions, but also "indications of interactive potentialities", which also have truth value, because can be falsified. The potentialities concern complex webs of counterfactual implications and possibilities, as a result, the emerging representational system will be highly complex, as well as contextual and environment bound. However, all covariance between the organism internal processes and environmental conditions are established incrementally by means of implicit predications that differentiate. The representational system is emergent in the construction of functional processes of interaction.

Now we can revisit the possibility of change and dynamic nature of representations in the practical context. We have said that the issue of change of representations surfaces most seriously when we consider 1) representations gain in precision 2) possibility of misrepresentation or inaccuracy of representations and 3) agent's own realization of an inaccuracy of their own representations. A model of representation modeled after the way implicit definitions define, instead of explicit definitions, can easily be rendered dynamic, because such a model allows: 1) Tightening the criteria, by means of adding further "axioms" to allow progressive differentiation and refinement, leading to increased knowledge. 2) There is room for misrepresentation, as in the fly – pebble example, to the extent that both go into the same equivalence class within the context of the presuppositions as axioms. 3) Error and agent detectable error is defined pragmatically, dynamically in contrast to implicit "corroboration". If the fly achieves to be fed, the representation was successful, and represented truly, if it fails, then it is falsified in a way which is transparent to the frog. (It is a further step whether

learning can occur in this case, depending on the frog's capacity to differentiate and add further criteria.)

Another example of the difference between implicit differentiations and explicit definitions would be the way "CAPTCHA" questions work. (See figure 2).

Figure 2. Example of a CAPTCHA prompt that relies on implicit reasoning

CAPTCHA are questions used to differentiate human users from bots since only the former can answer them. Arguably, the reason is that human agents work via implicit definitions in answering the questions. "Matching" the image of the cat with the images below requires not only comparing the images based on their actual properties, but also considering implicit assumptions regarding what the questioner might be assuming the answer to be. It involves considering which other image is relevant based on past experiences in other contexts. The criterion of success is not the same as the criterion of perfect correspondence. What the question is seeking is not a direct logical comparison between the images, but rather demands your understanding of context.

Another example of an implicit differentiation in a dynamic context is the range of operation of a thermostat.

Yet another example is implicit aspects of inter-personal encounters. When a child has been regularly neglected by his caregiver over a period of time, their presuppositions about their self-worth become so shaped as if "they are not worth caring for". This presupposition does not have to be explicitly represented in the child's head like a language of thought to be operative in his expectations and anticipations of interactions in the world, as if the child has such a maxim operating in the background (Bickhard, 1989). This explanation is the Interactivist alternative to the Freudian unconscious.

Furthermore, there is a use for actual encodings as internal and external stand-ins, which will be taken up with learning.

The distinction between implicit and explicit representations is important for this ontology of representation because it marks the rhetorical development from purely physical symbols as embodied structures to: 1) logical relations or rules 2) inferential transitions between representations or generated pragmatic results 3) functional emergence of normativity, better fit to function, not in terms of an objective omniscient observer but in terms of the agent's own anticipations of success and error.

2.3.4 Error, rules, and representations

Here is a toy example to demonstrate the relationship between error, rules, and our representations of what is correct or what is permissible, or possible, in other words, anticipations. Since representations are dynamic, the most illustrative examples are those

that show the incremental learning process, instead of taking a snapshot of the anticipations of what is correct and true at a point in time.

Imagine you go to a conference with a friend from work, and you stay at a hotel. You are given a keycard to open your room, which is on the twenty-second floor, while your friend is given a keycard for their room, which is on the fourteenth floor. The first time you get on the elevator together, you realize you cannot simply push the respective floor number buttons to get to your rooms, the elevator requires you to scan your individual keycards before allowing you to press the associated floor numbers with your individual cards, 22 and 14, respectively. In the upcoming days, you often get on the elevator together with your friend, and you observe him diligently scan his card before pressing the floor buttons, and getting to his room, successfully. There is one problem though, your friend scans his card also when he is trying to get to the first six floors, which are communal floors that do not require the card scan. Your friend of course, did not realize this, since scanning and not scanning ends up with the same result, he successfully gets to the communal floors and does not face an error or an "anomaly". If he were to keep trying new combinations of actions, inducing either heuristic variations (by reasoning something along the lines of "what if somebody who does not have a room tries to access the conference rooms" or curious to see "if the limit works for all floors" or observing others, etc.) or blind variations (like a child playing with the buttons), of scanning and not scanning his card, of pressing different buttons, despite the risk of looking stupid or being erroneous, he would have figured the exception to the rule, that the communal floors do not require the card scan. In not compromising the certitude, the guarantee of getting it right each time he tried, he goes through an extra

step, an unnecessary step, and his flow of action is not as rigorous as it could be, because his flow of action is more restrictive than it should be.

In this example, the friend's anticipation of possibilities of interaction are his representations. The very first time he got on the elevator and expected to simply press the button 14, to be taken up there, his representation was mistaken, and he detected his error. After the first failure, and his realization of the card scan requirement that led to his first success, his anticipation became fixed to represent that when he scans his keycard and presses the button the elevator takes him up or down the floors. The extent that the elevator does take the button-press-command and takes your friend as he intended to go, his anticipation (representation) is correct, it yields positive truth value. And yet his representations could have been much more precise, if he anticipated the results of more intricate actions, he would have known not only the truth value under one truth condition, but instead he would have known different truth conditions whereby he would achieve correct results. He would understand the conditions that yield truth and falsity (truth value). In any case, all he can know through interaction is the output generated by the rules; meanwhile the actual algorithm that operates the elevator, the explicit set of rules that define all the conditions of the elevator's operations are opaque to him. However, the rules generate reliable results, from the results he can extrapolate the conditions, the precision depending on the amount he varies his interactions. If he could interact with the elevator in every possible way, he would get a definitive list of possibilities of interaction, a topology of potentialities, without seeing the explicitly coded algorithm/rules. In other words, what the rule generates under all possible circumstances is equivalent to the rule, defines the rule, implicitly and not explicitly. The important caveat in this example is that the possibilities of interacting with the

elevator are finite and tractable, where one can exhaust them through trial and error, if not heuristics. This is often not possible in most of life's problems, as we will see in the chapter four and chapter five on frame problems, systematicity, and productivity.

The example is in principle the same with Hilbert's axiomatization allowing for different interpretations of the line segment under different physical frames of application. If the axiomatization is too restrictive, one can be confined to the Euclidian space, which is correct to the extent that it is. However, the real space of applicability is opaque, and we may not see it before we attempt to observe or interact with it. Applicability is a matter of discovery when frame of reference is not already given.

One may go both ways: one may go from applied instances, and infer the rules that generate the instances, like saying I start from a set of integer symbols {1, 2, 3, 4, 5, ...} and define the rule that generates the set as a system of counting numbers as, for instance "bijection".[23] The operation of assigning a counting number to a set of counted items implicitly assumes a bijective relation between the two sets (Zhuli, 2023). One does not need to know about the scope of bijective relation (and its difference from other mappings such as injective or surjective, see Figure 3) before learning to apply the rule, but one can infer from the application (counting) to the rule (bijection).

[23] Bijection is another name for one-to-one corresponding function. The function may be articulated as a rule and related to less strict conditions in the following way. Supposing that bijection concerns mapping of each element from a set A (domain) to a set B (codomain), the injective relation is the rule that elements in the codomain get mapped at most once; surjective relation is the rule that elements in the codomain get mapped at least once; and the bijective relation is the rule that both injective and surjective rules apply, so that elements in the codomain get mapped exactly once. Figure 3 illustrates this point.

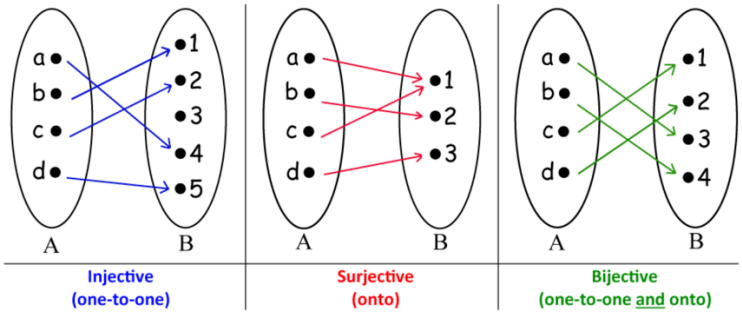

Figure 3. Difference between injective, surjective, and bijective relations as pairing conditions. Source: (Jenn, 2021) via https://calcworkshop.com/functions/bijection

Or one may start with the rules, the relations, and find out the instances that the rule applies, or the possibility it represents, for instance the equation $x^2 + y^2 = r^2$ represents a circle in Cartesian coordinates, because it relates the coordinates x and y with the radius r.

One may even have an incomplete set of working instances, like the example of our friend in the elevator, from which one extrapolates the rule, albeit strictly less than the actual rule, in this case.

All this said, there is a conceptual mismatch between formal axioms and dynamic presuppositions, which will be explicated in the next section.

CHAPTER 3

MATHEMATICAL DEFINITIONS VERSUS ONTOLOGY OF REPRESENTATIONS

Up to now we have talked about implicit definitions and their dynamic practical realization in the representational systems of animals as "dynamic implicit differentiations" as if they are of the same kind. However, one may object that it is a category mistake to apply the notion of implicit definition (implicit/explicit distinction) in logic and math to the notion of representations. Namely that definitions are types of logical operations while representations are ontological "entities"[24]. Ontology concerns what types of stuff exist in reality, what are their most general features and relations, what types of inferences are permissible. One may say it is not clear what kind of change to the ontological features of representation is argued for when we say the right model of representation should be based on implicit definition instead of explicit definition. Let us name the first objection: O1: logic versus ontology.

There is a second relevant objection. Saying that environmental properties are not directly represented, what is represented is the environment covarying functional organization of the organism, in terms of which environment is interacted with, is in effect collapsing properties with concepts. However, the objection would go, we cannot account for all the features of properties by means of the features of concepts. Let us

[24] "Entity" is used loosely to encompass both objects and processes. In fact, the preferred ontology of interactivism is process ontology whereby representations are ontological processes. However, this specification is not highlighted at this stage because the point I am making concerns the contrast between logic and ontology which should be provable before committing to the process assumption.

name the second objection: O2: properties versus concepts. The objections are related, and let's take them up in turn.

3.1 Logic versus ontology

Hilbert's axioms in his axiomatization of foundations of geometry are rigorous: the axioms are consistent with each other, independent of each other (no redundancies), and the axioms together form a complete system whereby one axiom may be proved in terms of the bulk of the others (with some assumptions). We say that the dynamic operation is analogous to the formal operation: the comparison is between how we interpret undefined primitives of an axiomatic system implicitly, to how we interpret differentiations in action implicitly. In what way can these crafted axioms be compared to the implicit differentiating of frogs for tongue flicking and eating? In what way is this analogical model literal? You will notice that this problem about mental representations not only concerns formal logic versus formal ontology, but also the relationship between formal and dynamic / pragmatic[25].

Part of the reason this problem comes up is that it was precisely to clarify the relation between logic and ontology that the received encoding notion of mental representation was developed. The logical, semantic, and syntactic properties of formal and natural languages were to be encoded by mental representations thereby bridging the content of thoughts, the content of propositions and (facts of) the world. When this view is rejected, the alternative view is burdened to address the same problem and the similar

[25] A similar distinction is knowing that versus knowing how, or propositional knowledge versus procedural knowledge.

problem of the relation between forms of judgement, forms of logical consequence and forms of practical /causal consequence.

The overview of the response goes as follows. Very simply and perhaps even trivially, *the logical operation of fixing an equivalence class based on implicit definition* would be *an aspect* of the ontology of representation. This aspect is meant to match the functional role of correspondence in the received view. It is meant to provide the veracity that a so-called correspondence relation provides, and further allow for the possibility of dynamism, changeability, partial correctness, fallibility, thereby accounting for phenomena such as misrepresentation, learning and agent detectable error. We have already talked about this functional role in the previous section.

To further explicate the possibility of *this aspect* (the logical operation of fixing an equivalence class based on implicit definition) as an ontological feature of representation, also shedding light on the ontological nature of representation will require pragmatic philosophy to systematically redefine the relationship between ontology, epistemology, and logic. Peircean pragmatism is a case in point where a revised understanding of the analytic as dynamic, fallible, and yet abstract and schematic is demonstrated. This revised understanding of the analytic allows to conceptually separate meaning and truth, such that meaning generation is defined as relational pattern systematized through habitual disposition and demarcated from truth as possible verifying instances.

It should be noted that, the question of the relation between forms of thinking, forms of logical consequence and forms of practical/causal consequence is already ill-formed if a shared “common form” that supposedly carries over between these domains is presupposed as the foundational condition of experiencing, interacting, perceiving or

justification. A shared “form” is not required for an agent -environment covariation or successful interaction to take place, because the interaction is not mediated through such a mechanism of commonality of forms. Instead, it more akin to the covariation and fit between the shape of a ship’s hull and the ocean’s features that allow the ship to float. Interaction is mediated through the functional organization flow of the organism. The functional flow may involve a sequence of sub-actions, a released hormone as preparation, a developed limb to execute the interaction, and their combination with relevant representational anticipations as branching pointers. There is no structural (nor syntactical) form shared between the domains in the first (direct) level of interaction. This is similar to Gibson’s position that we perceive surfaces and not forms in his ecological approach to visual perception.(Gibson, 1979/2015 p. 142) The appropriate level of description and language for the level of perception and interaction that we share with the animals should be anti-reductionist and biological, like Gibson points out. The proper vocabulary that describes the physical medium should be the ecological environment and not the physical reality, because the environment together with the possibilities of animal specific action it affords is directly meaningful, while the physical reality is not inherently meaningful in this way.

Thus “form” should be taken as a level of abstraction, a schema, an “image”, a “universal”, an abstract structure, a pattern, while a pragmatic instance of interaction is a particular. The issue is not that there is no commonality of “forms” between forms of thinking, logical consequences, and causal consequences. There most definitely is, otherwise abstraction would not be possible. Whereas granting shared, common forms does not explain the generation of these abstractions. The proper question of the relation

of forms between judgement, logic and actual experience is a version of an old one: "how did the universal *generate* the particular?"

The point we are making here is that the shared form is in the eye of the beholder, it is an abstraction of sorts, to provide part of an explanatory story. Philosophically, taking "the forms" as "given" is already invested in "the way of ideas". That said, so called "forms" may be shared in the abstracted, derivative levels that reflectively build (piggyback) on top of the first, direct, implicit level of interaction and affordances. (Gibson does not have much to say on theoretical activities of humans, unlike Peirce, Piaget and Bickhard.) This is an output, a schematization that the reflective consciousness achieves. At a sufficient level of abstraction an agent can reach a commonality of forms between their mental life and "properties" of the environment, however this is an achievement of after the fact translation and not the foundation of agent - environment interaction or knowledge. This is the capacity to make explicit, what was before implicit. And conversely what has become explicit may functionally become implicit via practice.[26]

There has always been the problem of the source of the relation between words and objects (and other versions of the problem of reference) in analytic philosophy. It is not that the path pursued in these endeavors is entirely incorrect. However, there is the missing link of "some sort of psycho-linguistic mediation" (Sellars, 1979), without which even the most basic logical or ontological formalizations lack their basis. For instance:

---

[26] This caveat will be revisited and explicated in the upcoming sections on "logic of action" and "explicit representations".

What is the referential link between abstract entities and their references?

What does it mean to say that the logical qualifier "'something' makes an indefinite or indeterminate reference to objects" in "something is a lion"?

or

What does it mean to say that variables "range" over objects?

Wilfrid Sellars raises this question with regards to the ontology of abstract entities, such as attributes, numbers, propositions, classes, definitions… etc. (1979). On the one hand, it seems like an obvious truth that there are these entities, on the other hand, their exact status and connection with reality, particularly with respect to determinate entities such as objects or persons is highly controversial. For instance, can one think of a world where all logical, abstract entities are reduced to non-abstract entities without loss of expression? The issue that he raises concerns the reality of logical classes, any named sortals including attributes, possibilities, definitions, not to mention normative psychological terms like affordances, habits, dispositions, anticipations, and the like. According to Sellars, for a quantifier like "some" (or "something is an x"), there is no escape from ultimately relying on modal notions [like necessity and possibility] in order to cash out its ontological commitments: any attempt at a rigorous paraphrase that appeals to enlisting particular determinate references of type x (Leo is a lion, the male lion in Barcelona zoo is a lion, Musti is a lion…etc.) is not even equivalent, let alone synonymous[27]; and any other paraphrased formalization that relies on the truth

[27] Is not equivalent (with the quantifier "some" or the expression "something is a lion") because in the above statement the expression "etc." is doing the work of topping up our limited list with the complete list of "all" determinate references. Then the question becomes, what grounds the interpretation of "etc."? "Etc." is as indeterminate or as inscrutable as "some". The formulation purports to account for the equivalence, but it is assuming what it purports to account for. It is an assumed equivalence and therefore not more legitimate than the reference "something" that it purports to paraphrase. At its base, it is

conditions for the usage of the quantifier is also modal[28]: i.e., "'Something is a lion' is true ≡ some statement which makes a determinate reference to an object and classifies it as a lion is true" (Sellars, 1979, p.7). The latter equivalence succeeds in tying indeterminate reference to determinate reference, but the achievement is not extensional, it is parasitic upon the rule-matrix of "holding of in context" (see footnote on Sellars' argument below). Sellars concludes that such formalizations either depend on the functionally derived modal characteristics of language (for instance extendibility) to fulfill their intended meaning (which is Sellars' position)[29], or else they must rely on an unspecified, unmentioned connection, that is implicitly presupposed but not explicitly argued for:

---

presuppositional, takes for granted what it purports to paraphrase. Therefore, fails to achieve an equivalence between determinate and indeterminate reference, instead paraphrases one indeterminate reference for another indeterminate reference. Thus the extensional (factual designation in the manner of (Sellars, 1956/1997)) equivalence remains indeterminate/ inscrutable.
The statement is not synonymous either, because 1) the point made above, lack of non-question-begging extensional equivalence 2) does not give the thrust of meaning that "something" gives, lack of intensional equivalence (L-equivalent or equivalent by virtue of semantical rules or equivalent in all possible worlds (Carnap, 1947/1988, p. 7)).

[28] The application/interpretation of the biconditional relies on the sortal expression being a linguistic utterance differentiated by the rules of application of those types of utterances within language. The argument in a nutshell, goes like this: for Sellars, neither extensions nor Leibniz type descriptions are sufficient to differentiate between untokened (vacuous) sortal types, like dragons and centaurs. So, we must be relying on an additional differentiation (/equivalence) criterion which is normative equivalence, based on equivalence of a conjunction of (linguistic) rules for holding in context for that type of sortal. Let p and b be untokened sortals (with empty extension sets) -of pawns and bishops in a "game which will never be played- suppose it to be chess". His suggested remaining criteria is "every context which holds of p, holds of b". Sellars substitutes "holds of" for "true of", for he takes "true of…", extensional as a special case of "holds of…", in context. While with extensional criteria, it is not possible to differentiate between p's and b's (because they are extensionally empty), if we stipulate Mp to represent conjunction of admissible pawn maneuvers and Mb to represent conjunction of admissible bishop maneuvers, we can differentiate between them based on "normative equivalence" or their rule-like expression, overcoming the difficulty with empty extensions. "'Let [(x)x is a --- ⊃ x be Mb'd]!' --- holds of bishops, but not of pawns." And vice versa. As a result if the general form of the rule-matrix holds of two sortals, they are normatively the same. (Sellars, 1979, p. 151)

[29] Extracting the modal, subjunctive properties from the functional and "rule-conforming" features of language is Sellars' position.

> But what is the alternative? It is, to my mind, a most puzzling one, though its puzzling aspects are quietly passed over; indeed, I would say swept under the rug by those who espouse it. For it amounts to nothing more nor less than the idea that the word 'something' has a connection, unmediated by determinate references, with all objects-and by 'connection,' as will become evident, I mean a genuine relation* to be captured by psycho-linguistic theory. (Sellars, 1979, p.8)
>
> *By 'genuine' relation I mean, roughly, descriptive relation. In this sense, neither 'or (p, q)' nor 'greater {9,7)' express relations. How, exactly, the distinction between 'descriptive' and its complement ('logical?') is to be drawn is part of the problem. (Sellars, 1979, p.8, footnote)

What applies to “something”, applies to the quantifier some (∃):

> I have been discussing the issue in terms of the word 'something.' But its bite remains when we transpose it into the language of the logicians, where
> Something is a tiger and is tame.
> becomes
> (∃x) x is a tiger and x is tame.
> The variable 'x' is said to range over objects. But it is not clear what it is for a variable to 'range' over objects. Is there a 'word-world' connection between variables and items in the extra-linguistic realm of stones and tigers? If so (and the answer must surely be 'yes') is this 'ranging'-which is clearly the counterpart of the indeterminate reference of 'something'-to be explicated in terms of determinate reference or is it to be taken as a basic mode of reference?
> . . . I have no objection whatever to treating, as a logician might do, the concept of indeterminate reference as an unanalyzed concept in semantical theory. But the task of explicating it confronts the philosophical logician as a challenge which should not be ignored and will not go away. (Sellars, 1979, p. 8)

We will come back to this issue on predication in the section “properties vs concepts”. Here we can note that the subject of properties vs concepts (section 3.3) is relevant to the subject of the current section, “logic versus ontology” through issues of predication and reference. To the extent that we take logic to be merely logical forms of truth, we face the problem of the ontological ground of reference of those forms. We agree with Sellars that the formalization begs explication, either through modality of language as he

suggests, or through some other psycho-linguistic connection which needs to be explicated.

It is precisely the function and contribution of pragmatic philosophy to provide this `psycho-linguistic` or `psycho-*logical*` link that occurs in hard knuckled encounter with reality. In this sense, pragmatic philosophy is complementary and prior to analytic philosophy. Analytic philosophy is already at the level of reflective consciousness, second level thinking, abstraction, and formalization. Pragmatism is the doctrine that there is a level before the explicit level, which makes the explicit level possible. That level concerns concepts like environment, interaction, habit, disposition, anticipation, verification. Pragmatism attempts to account for the things philosophy wants to account for such as truth, content, causal covariance in terms of these biological and psychological concepts. Pragmatism provides psycho-linguistic mediation that is left unexplained in orthodox analytic philosophy that establishes the referential link between abstract entities and their references. Reference is an achievement that is explained by pragmatism. Orthodox philosophy only begs the question in terms of commonality of forms, which is not a given but an achievement of mind. What orthodox formal philosophy talks about, is the part when representation becomes abstracted, explicit, denotational. Philosophy is constrained to talk at length about the genesis of these explicit, denotational, derivative representations in terms of modality, normativity, and other psychological concepts, as a result failing to make an established exposition of the mediating step that is psycho-logical.

The philosophical significance of the explication of (mental) representation is precisely to achieve the psycho-logical bridge.[30]

3.1.1 Pragmatic analyticity

Peirce's pragmatism provides a systematic response to the question of the relation between ontology and logic. We said, the logical operation of fixing an equivalence class based on implicit definition is *an aspect* of the ontology of representation. There is no discrepancy in taking a mathematical operation as an abstract principle for a logical operation and further taking a logical operation as an abstract principle for a more concrete operation in ontological philosophy. Laws of cause and emergence must abide by the more general laws of possibility and necessity. Peircean pragmatism, for instance, takes mathematics as the science of possibility and necessity in the most abstract form:

> Mathematics is the most abstract of all the sciences. For it makes no external observations, nor asserts anything as a real fact. When the mathematician deals with facts, they become for him mere "hypotheses"; for with their truth he refuses to concern himself. The whole science of mathematics is a science of hypotheses; so that nothing could be more completely abstracted from concrete reality. (Peirce, 1994, CP 3.428)

> This double assertion, that logic ought to draw upon mathematics for control of disputed principles, and second that ontological philosophy ought in like manner to draw upon logic, is a case under a general assertion which was made by Auguste Comte, namely, that the sciences may be arranged in a series with reference to the abstractness of their objects; and that each science draws regulating principles from those superior to it in abstractness, while drawing data for its inductions from the sciences inferior to it in abstractness. So far as the sciences can be arranged in such a scale, these relationships must hold good. For if anything is true of a whole genus of objects, this truth may be adopted as a principle in studying every species of that genus. While whatever is true of a

[30] Arguably, this implicit promise of achieving psycho-logical bridge was also the reason for the excitement about computationalist theories of representation and language of thought hypotheses in philosophy in the 1970s. The research program was directly relevant to the problem of reference and word – object correspondence.

> species will form a datum for the discovery of the wider truth which holds of the whole genus. (Peirce, 1994, CP 3.428)

Peirce's intuitive claim above that mathematics is the most abstract among branches of knowledge including various logics and philosophical ontology is counter to the traditional analytic philosophy which at some point attempted to reduce mathematics to logic[31]. The move was in the spirit of anti-psychologism, whereby Frege and later Russell aimed to save the objectivity of truth and justification from fallible reasoning capacities and contingent associations of individual thinkers.

Having also inherited anti-psychologism, the concept of the analytic, and the problematic of the source of "apriori" or "necessary" knowledge of mathematics from Kant, Peirce appealed to the concept of analytic (defined in his own way) to justify "an epistemic ideal world[32]" as a mathematical ("diagram-matical") system of ideas that shapes the way thinkers interpret the real world and a fallibilist epistemology, achieving similar ideals by means of a pragmatist philosophy. Anti-psychologism can be interpreted differently concerning epistemic and ontological anti-psychologism. For Peirce anti-psychologistic objectivity of truth did not lie in certainty in correct forms of true propositions, but rather in the cumulative verification and progressive convergence,

---

[31] Namely, Frege tried to cash out the notion of analytic truth in terms of the form of propositions, to arrive at form-*al* logical truths, by means of substituting nonlogical referring constants with their synonyms while keeping only the logical operators. This strategy did not work to reduce mathematics to logic due to problems with preserving concept meaning during substitution and synonymy, nevertheless the strategy was successful in terms of establishing the notion of a logical truth by means of the distinction between logical operators and non-logical expressions.
Frege's and our contemporary understanding of logical truth is meant to be objective, perspective less, neutral, and general. It is, however, the general position of this thesis that the notion of logical truth is only an idealization, it does not consider the reasoning activity of the subject. Peirce's ideas of neither logic nor mathematical thinking are idealizations. They are pragmatic and dynamic.

[32] We can understand Peirce's ideal world as the epistemic space of possibilities and necessities, akin to possible world semantics.

as well as an understanding of rationality in terms of critical reasoning. (Peirce, 1994, CP 5.108) His ontological anti-psychologism concerns separating meaning from psychology, in his hierarchy of abstraction: from mathematical relations to philosophy which concerns meaning and ontology to psychology which concerns positive realization.

Peirce's epistemic hierarchy of sciences starts with his observation that mathematics, albeit being fallible, is comparatively immune from dispute, because any mistakes, once suspected, are settled relatively quickly, a conclusion based on the history of sciences. This is because mathematical objects and relations are *creations* independent of reality, it is easy to repeat the same operations upon new instances, also taking up extreme cases as instances, to test the accuracy of the processes, if suspected to be faulty (Peirce, 3.426). Thus, instead of reducing mathematics to logic, Peirce offers mathematics to solve disputes of logical accuracy:

> [C]onsidering the immense amount of disputation there has always been concerning the doctrines of logic, and especially concerning those which would otherwise be applicable to settle disputes concerning the accuracy of reasonings in metaphysics, the safest way is to appeal for our logical principles to the science of mathematics, where error can only long go unexploded on condition of its not being suspected. (Peirce, 1994, CP 3.427)

More crucially for our subject, mathematics is more abstract than logic (as a branch of philosophy, along with metaphysics) is for Peirce, because mathematics is not a positive science, it is completely abstracted from concrete reality. Even philosophy is not so abstract, albeit not involving special experimentation and observation, and yet deals with reality in its most universal form. Logic is more abstract than ontology and metaphysics, and yet is still concerned with the truth of established beliefs and facts of human agents implied "in the supposition of an unlimited applicability of language" (Peirce, 3.428).

> Mathematics is not a positive science; for the mathematician holds himself free to say that A is B or that A is not B, the only obligation upon him being, that as long as he says A is B, he is to hold to it, consistently. But logic begins to be a positive science; since there are some things in regard to which the logician is not free to suppose that they are or are not; but acknowledges a compulsion upon him to assert the one and deny the other. Thus, the logician is forced by positive observation to admit that there is such a thing as doubt, that some propositions are false, etc. But with this compulsion comes a corresponding responsibility upon him not to admit anything which he is not forced to admit. (Peirce, 1994, 3.428)

> Logic may be defined as the science of the laws of the stable establishment of beliefs. Then, *exact* logic will be that doctrine of the conditions of establishment of stable belief which rests upon perfectly undoubted observations and upon mathematical, that is, upon *diagrammatical,* or, *iconic,* thought. (Peirce, 1994, 3.428)

For Peirce, diagrammatical thinking is mathematical thinking, and it particularly deals with hypotheticals, hypotheses. The link from a hypothetical axiom to a theorem is conditional. Diagrammatical thinking is concerned with relations only, that hold between hypotheticals, essential or possible variations from premises toward conclusions, relations of necessity or possibility that can be derived from schemas or diagrams. The science of mathematics is not concerned with the applicability of these schemas, and mathematical (diagrammatical) thinking does not guarantee applicability. Indeed, the application of context cannot be captured within the formalization since diagrammatical thinking concerns precisely the abstraction from it.

Diagrammatical thinking captures what we would currently take to be "analytic" in terms of the study of valid reasoning, yet Peirce stresses the exploratory nature of such thinking activity, regardless of its abstract form, rather than taking "the analytic" as infallible body of knowledge. Diagrammatical thinking is fallible and revisable, albeit in different ways than positive knowledge. It also allows novel truths to be uncovered, and thus is creative. In these respects, the analyticity of diagrammatical/mathematical

thinking may seem to keep the cake and eat it, too. It's justification, however, lies in the synoptic consistency and explanatory power of Peircean philosophy.

Peirce's understanding of logic instead is concerned with emergent logical structures. Logic is an output of, derivative from diagrammatical (mathematical) thinking plus "undoubted observations", which is more general, abstract way of understanding rational reasoning. Accordingly, Peirce's understanding of logic is broader and different than our current sterile understanding of classical logic as the study of valid (logically true) forms of argument within a given formal or informal language and the associated model-theoretic semantics to establish the domain of interpretation. Contemporary understanding of classical logic seems to coincide with Peirce's understanding of mathematics in terms of their subject matter of valid logical forms and analyticity[33], but there are important differences. Peirce's notion of mathematics involves relations of ideas to be dynamically uncovered whereby novel (mathematical) truths may be uncovered. Furthermore, Peirce's pragmatic analyticity is not threatened by the dynamic and fallible activity of the thinking subject. This is because his warrant of reliability of truth and justification does not lie in *fixed forms of truth*. Truth is instead, derivative from the direct encounter with reality in lived experience or scientific experiment; these direct interactive encounters have the power to falsify the thinker's ontogenic system of ideas or the validity of their diagrammatical thinking. Engaging in diagrammatical thinking involves a dynamic *encounter* with a diagram or a schema, from which novelty or revision may arise.

---

[33] Albeit 70 years of disputes about how to distinguish analytic from synthetic knowledge within the analytic tradition.

On the other hand, if we take logic as accounting for the ground of truth via forms as in the analytic tradition, the truth of logical forms becomes a metaphysical matter and a reification of sorts. If, however, pragmatic actual and counterfactual possibility and necessity is prior to truth in our reasoning, and mathematics is the investigation into it, logical truths become derivable from this kind of mathematical (diagrammatic) reasoning.

3.1.2 Diagrammatical reasoning

Mathematics is the science of what Peirce calls diagram-matical or schematic reasoning, whereby one draws out the realization of an argument in schematic form to test the validity of its conclusions. Significant relations as minimal conditions are represented, and the conclusion is observed to be necessary or not. This is essentially what we are doing in algebra, according to Peirce.

Peirce gives the story of a kingdom with provinces and a river running through it as a simple demonstration. In his example, one cannot show if an inference is necessary without constructing a diagram with systematically repeated signs and symbols subject to rules that correspond to the relations associated with the entities (1994, CP 3.418).[34] A

[34] "Consider any argument concerning the validity of which a person might conceivably entertain for a moment some doubt. For instance, let the premiss be that from either of two provinces of a certain kingdom it is possible to proceed to any province by floating down the only river the kingdom contains, combined with a land-journey within the boundaries of one province; and let the conclusion be that the river, after touching every province in the kingdom, must again meet the one which it first left. Now, in order to show that this inference is (or that it is not) absolutely necessary, it is requisite to have something analogous to a diagram with different series of parts, the parts of each series being evidently related as those provinces are said to be, while in the different series something corresponding to the course of the river has all the essential variations possible; and this diagram must be so contrived that it is easy to examine it and find out whether the course of the river is in truth in every case such as is here proposed to be inferred. Such a diagram has got to be either auditory or visual, the parts being separated in the one case in time, in the other in space. But in order completely to exhibit the analogue of the conditions of the argument under examination, it will be necessary to use signs or symbols repeated in different places and in different juxtapositions, these signs being subject to certain "rules," that is, certain general relations

toy example in the same vein would be having to solve an "identify the blood relations" puzzle by means of a chart. Consider:

> Premise: "A girl introduced a boy as the son of the daughter of the father of her uncle."
>
> Alternative conclusions to be tested: "The boy is the girl's (1) brother (2) son (3) uncle (4) cousin."

Once all necessary and possible relations, as well as different signs for the different types of entities are laid out in terms of a diagram or a schema, possible variations become explicit, and the solution becomes "observable" in Peircean terms. The more systematic the schematization, in terms of different symbols to explicitly represent known differentiations of types of entities female, male, cousin relation, sibling relation, parent-child relation, the more explicit[35] the solution becomes.

In Figure 3, we see a schematic sketch of possibilities. The relations specified in red denote possibilities that become explicit for a thinker once the problem is schematized in blue.

---

associated with them by the mind. Such a method of forming a diagram is called **algebra**. All speech is but such an algebra, the repeated signs being the words, which have relations by virtue of the meanings associated with them. What is commonly called **logical algebra** differs from other formal logic only in using the same formal method with greater freedom." (Peirce, 1994, CP 3.418)

[35] We will come back to the differences between explicit and implicit representations in following chapters.

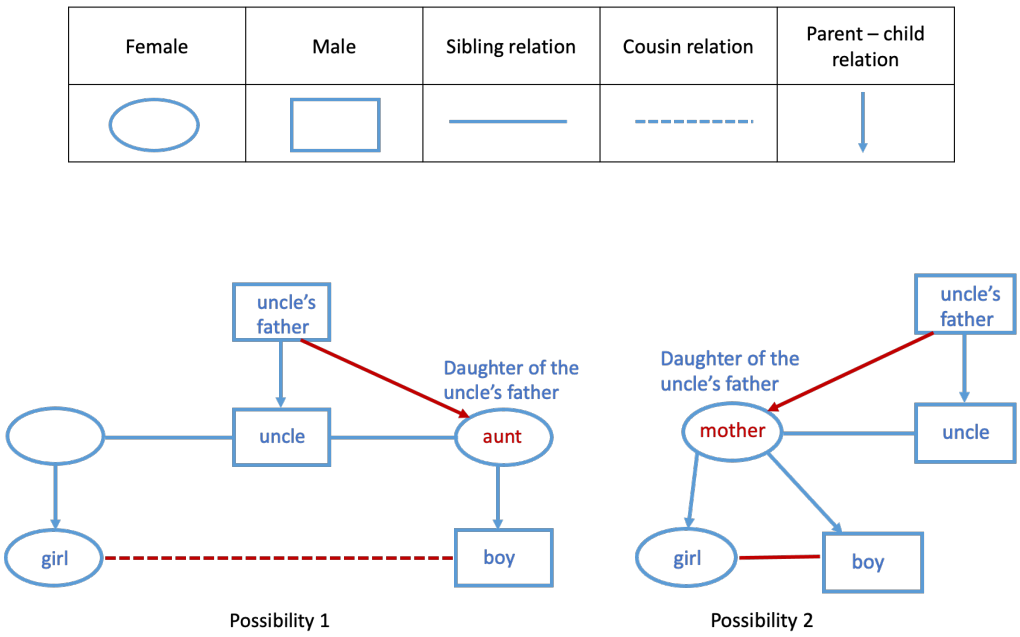

Figure 4. The diagrammatic solution of a word puzzle

The nature of mathematical reasonings also point out the need to differentiate between verification, truth and meaning. Mathematics is the study of necessary deductions from hypothetically possible postulates. Even though mathematical truths are necessary truths, mathematics does not provide certitude, mathematics is not infallible in two ways: 1) the deductive process may err, one may always make a cognitive mistake in following a rule, as in 1 + 1 =3; 2) we may be mistaken about the applicability of the mathematical structures to reality: "Mathematics does not provide certitude about the world. Peirce here stresses that he does not mean that if pure hypotheses were to be true of an actual state of affairs, then the reasoning would cease to be necessary. Rather, one could never know apodictically that these hypotheses were true of an actual state of things." (Rosenthal, 1994, p. 23) On the contrary, the ideal of the logical truth model is to achieve certitude so that reasoning would cease to be necessary.

Furthermore, in mathematical reasoning, there is still room to derive and discover new relations, new necessities, new mathematical truths. Peirce accounts for the ground of this possibility by means of diagrammatic reasoning whereby we interact with the schematization of a mathematical problem. Peirce defines a schema as that which is capable of being interacted with ("observed") *as a particular object* as well as being *a general* (a universal):

> The Diagram remains in the field of perception or "imagination" and so the Iconic Diagram and its initial Symbolic interpretant taken together constitute what we shall not too much wrench Kant's term in calling a schema, which is on the one side an object capable of being observed, while on the other side it is a general. (as cited in Rosenthal 1994, pg. 26)

Here we arrive at Peirce's pragmatic conception of the analytic – synthetic distinction. Peirce defines the analytic as the necessary inferential relation between a hypothesis and its conclusion, instead of the relation between the subject and predicate of a statement. Peirce's understanding of analyticity is not synonymy or equivalence, but rather "containment": "Thus, a mathematical system is a system of necessarily related meanings such that the meanings set by the starting postulates necessarily contain that which they have the power to generate." (Rosenthal, 1994, pg.23) Analyticity is an imposed relational structure while synthetic is the experiential part. However, even within the analytic enterprise of mathematics, it is possible to realize novel relational structures through the construction of schemas, or schematic reasoning. This point shows that the relational structures, even though necessary, are not static, they change, and there is a dialectical relationship between the agent and the necessary truths in theoretical or practical domains. In fact, if change and development is possible in the analytical domain, the same possibility would apply to the practical domain.

3.1.3 Representation vs definition

In studying the phenomenon of mental representation philosophically, the hierarchy of abstractness between mathematics, logic and ontology comes up in recursive ways. In studying mental representation philosophically, we are attempting to study the knowledge and reasoning structures whereby we grasp and make sense of the world – the "psycho-linguistic link". As such, our conception of the representing faculties is subject to discussions of accuracy and consistency both internally and externally.

First, there is the issue of metaphysical possibility, our hypothesis that "the nature of mental representing as implicit defining" must be metaphysically possible. This point concerns "what" question and biological vocabulary, and in the interactivist account, it is explicated in terms of being emergent in the functional organization of the organism. Relevant vocabulary is organizational flow and implicit differentiations as pointers. The above point about "aspect" is meant to address that level, but it is trivial without other levels.

Secondly, the account must meet the criterion of epistemic possibility and sufficiency, emergent in environment – organism interaction. This concerns psychological vocabulary, such as organism – environment match, possibility of origin and development, affordances, achievement of extensional correspondence...and so on. The epistemic nature of mental representing as implicit defining must be epistemically explanatory.

Thirdly, the account must be reflexively or internally consistent, that the implicit way of representing must match and explicate the internal logical and semantic accuracy of mental representing. This point concerns issues such as relation among representations, achievement of intensional correspondence, relational implicit

definitions, and their adequacy in achieving the meaningfulness, accuracy and expressiveness that we attribute to correspondences. The three levels operate in relevant domains of rules, subject to different norms/normativities: organism/ survival norms, epistemic norms, and semantic norms.

Taking representation as the psycho-linguistic (or psycho-logical) mediator between abstract entities (content) and experiential encounter imposes both internal and external questions to the consistency of the notion of mental representation. This was the case also in the encodingist paradigm, however since that paradigm takes "common forms" shared by judgements, propositions, and facts (or more sophisticatedly between inferential consequence, logical consequence and causal consequence), it takes for granted and does not problematize the source of the forms. Pragmatism rejects this and offers a dialectical story of origin between organism, environment and abstract content.

Peirce was a philosopher interested in mathematics, mathematical reasoning and semiotics (as study of meaning and representing), so his hierarchical framework is relevant to the recursive complexity. According to Peirce, there is a difference between independent, (indeterminately vague and continuous) reality, the real world, and the perceived world which is shaped by the system of ideas and meanings through which we grasp the real world, "what there is". On the one hand, the system of ideas and meanings make up a bigger set, as they include not only what is actual, but also what is possible and conceivable, an ideal world of possibilities. Peirce says, "the sensible world is but a

fragment of the ideal world." What is consistently thinkable, conceivably possible limit what might actually be anticipated[36] to be possible:

> [world of] perceptual facts." Such a world is a consistent system of facts rigorously obeying the laws of non-contradiction and excluded middle, for "Dichotomy rules the ideal world," the process of sensible experience to locate its facts in the world of ideas." such a grasping of the sensible world in terms a system of ideas is of the very essence of the sensible world. (Rosenthal, 1994, p.3)

For Peirce the epistemic ideal world (whose nature is a diagrammatical – relational system of meaning) is prior because it concerns relations of possibility, from which actuality evolves. Concerning qualities, basic even though actual sensation is a requisite for the apprehension of a quality, no sensation (as actual) is "requisite for the possibility which is the being of quality". Possibilities of what could happen or what would necessarily happen are what gives qualitative character to reality, and "It is impossible to hold consistently that a quality only exists when it actually inheres in a body", against nominalism. One reason for this that he points out is that in rejecting "laws" [of nature] as nugatory and fictitious, the nominalist position is relinquishing the license to talk meaningfully about an "instant", and therefore "uniformity". "If no law subsists other than an expression of actual facts, the future is entirely indeterminate and so is general to the highest degree" (Peirce, 1994, CP 1.422) Thus the prior relations of possibility[37] gives the determinate character or qualities to experiential reality.

---

[36] Remember that according to Interactivism, anticipation (of possibilities of action) is the content of representation. This is in line with saying that the sensible world is grasped by the agent's system of ideas, the agent's ideal world of what is anticipated and what is possible in any particular instance as well as in general.

[37] Necessity is derivable from possibility within a system of reference. This is a way to define systematicity. We will revisit this subject in Chapter 5: framing, unity and systematicity.

On the other hand, the actual, perceptual world is independent of our conceptions of it, because we can always be falsified and proved wrong in our direct encounters with the actual world, whereby "the ideal world" of our system of ideas (of what is necessary, possible and conceivable) would also change. This is a consistent position within Peirce's pragmatist framework because the way we get to know about the world is through experimental verification instead of spectating. Verification as experimental encounter is direct interaction. This is a very similar position to the interactivist account highlighted before, where the verification and falsification as contact is ontologically separated from the disposition and anticipation as content. This allows the external world to be partly opaque in all its properties and yet scientifically real.

Another way to make the same point is to distinguish between a theory of truth and a theory of meaning as Peirce and Bickhard does. Verificationalism provides a theory of truth, while disposition, anticipation, provide theory of meaning or content. In explicit definition, truth and meaning are artificially converged, in implicit definition, they are still separate as in their genesis. Truth lies with experiential encounter with ontological reality, meaning lies with what is epistemically possible, within the space of reasons. Explicit definition is static because of fixing truth with meaning at a moment in time. Implicit definition allows dynamism because truth and meaning are separate and changeable in time: meaning (rules of generation) follow truth value (the result(s) of verification/falsification) and anticipations for true results follow from meaning as rules of generation, diachronically.

Scientific realism concerns realism with regards to the independent existence of the real world:

> Peirce, in rejecting the role of humans as spectators, in understanding experience as a unity of interaction between humans and that facticity that gives itself within experience, can hold at once that the real world is the perceived world, that the real world has an independence from mind, and yet that the perceived world is partially dependent upon the noetic act and is thus relative in its nature to the mind. The supposed incompatibility of these three characteristics of the relation of thought to the real world stems from failure to radically and once and for all reject the presuppositions of a spectator theory of knowledge. (Rosenthal, 1994, 2)

What Peirce calls the spectator theory[38] of knowledge overlaps with the correspondence theories of knowledge and encodingist models of representation. The pragmatist understanding takes representational (interpretational) accuracy as stemming from a disposition as a rule, utilizing the relation of *generation* instead of copying. An example of a relation of generation would be "a mathematical rule generating[39] a number series or, as Peirce often explicitly develops it, of a continuity containing an infinite number of possible cuts" (Rosenthal, 1994, 34), which also apply to implicit defining that generates instances that satisfy its conditions.

The relational structures that we impose as meanings emerge from behavioral response patterns in interaction.

> For Peirce, meanings are to be understood as logical structures, not as psychological or biological facts. Peirce does not since it does not adequately ground the issue in its epis- want to give meaning an existence independent of purpose, yet he does not want to reduce meaning to the categories of psychology

---

[38] The contrast is between a strictly dynamic point of view versus an external semi-gods eye view. According to pragmatism, the spectator/ god's eye point of view is parasitic upon and derivative from the dynamic point of view which is fundamental.

[39] One significance of a generation rule is that stands out in this example is that the rule itself cannot be understood independently of the output which the rule generates. For instance, we are defining the generational rule by its output, say the number series "it" generates, such as "66 69 72 75 …". To the extent that we understand the relational structure of the output (its commonality) we understand the rule that generates it. So, the ground of rule-based generation can only be studied in terms of the aspects of its outputs. This point will be revisited in further sections when we talk about "the nature of disposition" as the nature of the ground of meaning.

> or biology. Meanings are to be understood, for Peirce as relational structures emerging from behavioral patterns, as emerging from the lived-through response of the human organism to that universe with which it is in interaction. Or, in other terms, human behavior is meaningful behavior, and it is in behavior that the relational patterns that constitute conceptual meaning are rooted. What, however, is meaning as a relational pattern? (Rosenthal, 1994)

When we pose the question in terms of "what is meaning as a relational pattern?" the problem seems enigmatic philosophizing. What is being explained however, is a many stepped, complex construction, and it involves possibilities and constraints on biological, psychological, semantic/logical/relational levels. However, it is constrained by the progression: from instances of interaction (behavior) to habit (repeated interaction), to disposition (acquired learning) to expectation (anticipation and interpretation of possibilities), to representational content as relational patterns. When we are talking about content, we can only refer to the momentary snapshot schematized as relational patterns, as an abstraction. The progression would shape the course of following interactions, formation of habits, further shaping expectations and conceived topologies of interactive possibilities and potentialities. Implicit differentiators fixing an equivalence class is the operative relation that limits and constrains.

3.1.4 Meaning as relational patterns

If we accept that the operation of fixing (defining) an equivalence class by means of an implicit definition is an operational aspect of representation, we may try to understand where this aspect fits in the overall ontology of mental representations. Representation has other aspects apart from the structure of logical operation that generates conclusions.

The tension between definitions and representations in the Interactivist account above is comparable to the problem of the two facets of *habit* in Peirce: the epistemic

and the ontological dimensions of habitual meaning. The difference maps onto the applicability of meaning versus its origination, generation.

> [Meanings embodied in the conceptual scheme are] built up in light of past experience. They are drawn from the empirical situation, although the relation among the meanings is stable apart from any particular instance of fact. The origin of our analytic structures, then, is empirical, pragmatic, functional. This genesis of meanings from the context of experience is in no way analogous to the logical reducibility of meanings to experience. The first answers the question as to why we create the meanings we do; the second answers the question as to what a meaning is. Meaning, qua created structure, contains no truth claim as to applicability in experience. Though for pragmatic reasons we must create or abductively fixate meanings with workable applications in the ongoing course of experience, a meaning itself is a deductive system applying to a hypothetical state of affairs, the implications of which we can know about since we create it. (Rosenthal, 1994)

According to Rosenthal, Peirce's pragmatism does not reject the analytic – synthetic distinction but develops his own interpretation of analyticity. Analyticity concerns the deductive system whereby certain conclusions follow from the premises. Mathematics is analytic whereby an arbitrary hypothesis is presupposed, without regard to its actual application, and possible conclusions that deductively follow are run through. An argument or logical structure is created that may or may not apply to empirical experience. According to Peirce analyticity involves containment (inclusion/ subsumption) relation (and not identity).

> The inherent relatedness to experience of our empirical meanings…arise through the fixation of a set of experienced relationships unified by habit as a rule of organization. The fixation intended corresponds most closely to the creative process that Peirce calls "abduction," though what are here fixed by such creative activities or abductive processes are not empirical hypothesis asserting the applicability of meanings but the very structure of the meanings themselves. As Peirce states, "An abduction is Originary in respect to being the only kind of argument which starts a new idea." (Rosenthal, 1994, p. 39)

However, Peirce's pragmatic maxim mentioned here is not an operationalism:

> Israel Scheffler's assumption that Peirce's pragmatic maxim is similar to Bridgman's operational definition, in which meaning is reducible to verifying instances and in which the "more than" disappears, underlies his claim that Peirce must deny the validity of theoretical structures in science. Perceptual objects are "theory-laden" in the sense that they are constituted via the meanings through which we perceive our world and that we have brought to our world. Perceptual meanings are theoretical in the sense that they are the explanations or interpretations by which that which is there reveals itself to us. Genetically, our meanings arise through the cumulative effect of past experience and the abductive, creative fixation, within the ongoing course of experience, of dispositionally organized relationships among experiences. But at any point in the knowledge process, the meaning logically contains all that it has creatively fixated or, conversely, all that it now has the power or potential to generate. (Rosenthal, 1994, p. 39)

Another way of articulating the above is that sequences and possibilities of actions must also follow analogous criteria of logic: they must not be inconsistent with one another (one hand must not undo what the other hand does), independence is preferred if not mandated (no redundancies would be selected for), and their flow must necessitate each other as part of a complete flow. In fact, logic must derive from the possibilities of action and not vice versa. Possibilities of actions are systematic, what actions would follow from a set of previous actions is “contained”.

3.1.5 Pragmatic realism and epistemology

One objection would be with regards to reality and pragmatic understanding of realism. If the truth of our representations is established through interactive experience, and if there is no direct correspondence between the knower’s representation and the external reality, does that lead to an idealism?

The question partly depends on what we mean by idealism. If we take idealism as the ontological constitution and dependence of reality, as nature itself on mind, or solipsism in short, the answer is no, this sort of mind and interaction mediated

correspondence does not lead to an ontological idealism. Nevertheless, this account admits the mind dependence of our understanding of nature, via the mind constructed concepts which draw out the limits of conceivability, that Rescher (1991) calls "conceptual idealism", as opposed to "causal idealism":

> [Conceptual idealism] maintains that an adequate descriptive characterization of physical ("material") reality must make implicit reference to mental operations - that some commerce with mental characteristics and operations is involved in explanatory exposition of what is at issue "the real world." The central thesis of this position is that the mind is responsible for nature-as-we-understand-it, through making not nature itself, to be sure, but rather the mode-and-manner determining categories in whose terms we conceive of it. On its approach, the constitutive role of mind is accordingly to be thought of neither in ontological nor in causal terms, but hermeneutically by way of concept-explication.

Epistemology is dependent on conceptual framework and its logical structures imposed by the agent on to the world interpreted. These are unbounded yet finitary, while the reality in its level of detail is both unbounded and infinitely rich.

Rescher also makes a distinction between the capacity of "modeling" versus "conceptualization", the latter of which is uniquely mental and necessarily mind-involving, while the former can also be achieved by computation. (Rescher equates modeling with representing, which is different from our usage here. What he means modeling is a mere isomorphism) What makes conceptualization special is firstly the thinker's capability to bring in ideas to anything considered. This bringing-in involves fundamental reflexivity "in the mind's view of its own role in the scheme of things.", and "fundamentally relational" properties of all our conceptualizations, involve a reference to the mental operations within their very meaning. Rescher avoids a circularity by means of "a cognitive coordination of mentalistic and materialistic concepts…with respect to …their mutual interdependencies" (p.515). There would have

been a true circularity if a “causal idealism” were involved, however there is no circularity with “conceptual idealism”:

> "Surely roses would smell just as sweet in a mind-denuded world!" Well. . . yes and no. Agreed: the absence of minds would not *change* roses. But rose-fragrance and sweetness - and even the size and shape of roses - are features whose character hinges on such mind-invoking operations as smelling, scanning, measuring, and the like. For something really to be a rose it must, unavoidably, have various capacities to evoke mental responses. It must admit of identification, specification, classification, and property attribution, and these, by their very nature, are all mental operations. (Rescher, 1991, p.518)

This sort of conceptual idealism does not lead to a solipsist or voluntaristic, nature-independent coherence. Any notion of reality independent coherence is ruled out by the dependence of representations on direct contact with reality within experience. Interaction is direct contact, and direct contact imposes limits on representations in terms of falsification, independent of interpretation. Not everything is dependent on interpretation. The pragmatic limit is also an ontological and epistemological limit. (Magoo cases are acceptable, but they are not substantially possible long term).

A legitimate kind of correspondence (covariance with environmental or contextual conditions) can be expressed in terms of the structure of potential interactions:

> The basic model is highly compatible with a correspondence theory of truth and an evolutionary theory of growth of knowledge (Popper 1965, Campbell 1974). The correspondence, however, is not with the world-in-itself but rather with the structure of potential interactions (Rescher 1973, but with a distinctly different conclusion with respect to the nature of truth). (Bickhard, 1980, p. 214)

Being action based, it is more accurate to measure correspondence not with degree of fidelity and correctness (i.e., truth) but with amount of and criteria of error (i.e., falsity).

Error criteria is what defines different areas of human knowledge. Truth and falsity arise already at the level of general action systems:

> Truth and falsity arise with regards to the accuracy of representations with regards to what they represent. In the Interactivist model, truth and falsity arise at the level of general action systems and not language (since language is not taken to be the fundamental level of representation) More specifically, the situation and world images may be in error. (Bickhard, 1980, p. 112)

The idea of correspondence mediated by interaction derived concepts brings up Hume's problem: justification of habit (and the justification of laws of nature) is problematic without any independent foundations. Let us tackle this problem in the next section by means of the link between logical and causal consequence, or the link between logic of inference and the logic of action.

3.2 Logic of action and modal inference

The purpose of this section is to explicate the derivation of inference and epistemology from the flow and logic of interaction, interactive procedures. In pragmatist philosophy "meaning as habit" is supposed to link logic of action or behavior to the logic of inference and meaning. Representation as "interactive anticipations of action possibilities" also seem to allude to the parallel between interactive and logical consequence. But it is obvious in both cases that there is no direct overlap. So how is it supposed to play out?

In the history of philosophy, one direct counter alternative to correspondence and representationalist accounts of intentionality was "meaning as use", or inferentialist role semantics accounts. These accounts also take interactive properties such as dispositions, functions, and pragmatics as prior to meaning. We need to understand the tools they

used, namely modal properties, to be able to understand better the new interpretation interactivism brings to the modal properties of procedures, as well as the link between interactive, procedural; and declarative and logical consequence.

Encodingism is a kind of representationalism. Representationalism is "the view that representation comes first, then inference, then use." (Millikan, 2005, p. 69) Interactivism joins camp with a host of other positions (that may or may not call themselves pragmatist) such as Sellars's, Brandom's, Millikan's in arguing against representationalism.

3.2.1 Meaning as use

The alternatives to representationalism are different flavors of explaining meaning and intentionality in terms of "use first" theories. Such theories take into consideration the interpreter or "consumer" (Millikan, 1989, p. 286) of representations that take the representations as representing as inseparable from the semantics of representations. Therefore, the order is reversed, from pragmatic or functional use to inference and semantics. Semantics of representations are understood as "the involvement of truth or *satisfaction* conditions" (Millikan, 2005, p. 70). The full emergence of satisfaction conditions is owed to ways of performing: procedures or mechanisms whereby the function or pragmatics are realized. In the broad category of "use", a difference between an act as a procedure and its result as a performance is recognized.

The interpretations of what I cluster together here as "function or pragmatics" differ greatly between accounts. Here I am clustering them for the purposes of generalizing over their commonality of utilizing modal vocabulary within the history of analytic philosophy. To briefly highlight the extent of differences; for example, for

Millikan, function is what she calls “proper” or “stabilizing” function, preserved through social, learning, or evolutionary histories of active selection and performance (1989, p. 286). She takes the concepts of “survival value” and “picturing” from Sellars. From use emerges truth or satisfaction conditions, and from truth or satisfaction conditions emerges complete representations with intentionality. Millikan puts the semantic progression from the functions of inferences to engaging in intentional attitudes (such as joint attention or agreement) to intentional content:

> Briefly put, there is no such thing as intentionality without attitude. Participating in inference processes by which new descriptive and directive representations are formed is a central way in which human intentional attitudes are employed, so the intentionality of [human intentional] attitudes and their content is a function, in part, of inferential patterns…On the other hand, I have claimed, there is also plenty of intentionality prior to that of the intentional attitudes, both in the perceptions of animals and humans and in simpler messenger systems that abound in the body. (Millikan, 2005, p. 70)

Thus, the scope of intentionality for Millikan is wider, also involving the intentionality of non-human animals and sub-personal processes. For Brandom, on the other hand, intentionality is derivative from pragmatics, which consists of the collective linguistic commitments. The meanings that linguistic agents understand to produce and to interpret, commit the agents to inferential reason giving and asking practices. Meanings are based on abiding by the linguistic rules of language games, that agents are engaged in. The conceptual content is derivative from the inferential relations involved with each concept by means of the linguistic rules (lawlike occurrences) which the users commit themselves to in engaging in these practices. The “strong inferentialist” position of Brandom is differentiated from Sellars in the sense that, for Brandom, it is norms all the way down, inferentialist relations are sufficient to determine meaning (of concepts, conceptual roles in inferential relations) altogether, while for weaker versions of

inferentialism, this is merely a necessary criterion of concepts, and not sufficient (Brandom 2001, p. 600).

Both Millikan and Brandom are inspired by Sellars, who tried to incorporate two Wittgensteinian themes: the "language games" metaphor from *Philosophical Investigations* (1953), which was an inspiration for inferential role semantics, and the causal-order relation of "picturing" (as representing), that is supposed to hold between word and world, in his earlier *Tractatus Logico-Philosophicus* (1922). Sellars' position was that, learning the meta-linguistic rules involved learning to use concepts according to their inferential roles as they are used in a specific language. He talked about a "schematic world story" (Sellars, 1979, p. 105) within language about the world and "fantastically complex rules of projection"(Sellars, 1979, p. 118) whereby formal and material inference rules dictating a system of sentence to sentence transitions mapped to the world represented. Despite talking about mapping and schemas, the mapping lied not in the structural correspondence between world and word as in the representationalist (or encodingist) theories, but in the bindingness of norms of language for language users. Thus, the patterns formed by such a system were not merely patterns of fact, but rather patterns of engagement, prescriptive and evaluative regularities of attitudes of users. Distinguished between regularities in behavior and "obeying a rule"[40],

---

[40] Rule obeying is a concept of Wittgenstein from *Philosophical Investigations (1953)*, for instance: "206. Following a rule is analogous to obeying an order. We are trained to do so; we react to an order in a particular way. But what if one person reacts in one way and another in another to the order and the training? Which one is right?" (p.82e)

And there is a stability to the rule obeying, as he states in: "The common behavior of mankind is the system of reference by means of which we interpret an unknown language." (p.82e); "240. Disputes do not break out (among mathematicians, say) over the question whether a rule has been obeyed or not. People don't come to blows over it, for example. That is part of the framework on which the working of our language is based (for example, in giving descriptions)." (p. 88e); due to the agreement in "form of

Sellars derived intentionality of content from the dispositions of users to follow the linguistic and meta-linguistic rules:

> [L]earning to use a language is coming to do A in C, A' in C', etc., *because* of a system of "moves" to which these acts belong, while yet denying that learning to use a language is coming to do A in C, A' in C', etc., *with the intention of realizing* a system of moves. In short, what we need is a distinction between "pattern governed" and "rule obeying" behavior, the latter being a more complex phenomenon which involves, but is not to be identified with, the former. Rule obeying behavior contains, in some sense, both a game and a metagame, the latter being the game in which belong the rules obeyed in playing the former game as a piece of rule obeying behavior. (Sellars, 1963, p. 328)

According to Millikan (2005, p. 63), Sellars' position is limited to dispositions to behave in certain ways, i.e., dispositions to allow certain rules in the language game and dispositions to make certain moves in the language games, and so on. The realm of dispositions forms a factual level of reality ("the space of reasons") which can be studied scientifically as they are, however "normative rules, for Sellars, are not translatable into non-normative terms. Accepting a normative rule is not believing a fact but tending to be motivated in a certain way" (Millikan, 2005, p. 62). This is where Brandom has a stronger position on inferentialism than Sellars, because Brandom claims there is a realm of meaning of conceptual content whose meaning is independent of dispositions.[41] So, there is a realm of "conceptual norms", apart from "dispositional norms" in Brandom's

---

life" in people's "agree[ment] in the language they use" (p. 88e); "I should have said: This is how it strikes me. When I obey a rule, I do not choose. I obey the rule blindly." (p. 85e)

[41] This thesis agrees with the point regarding dispositions and content of concepts, but it is doubtful that the criticism applies to Sellars' full position. The analogy to the game moves would lose world-word ties if we only interpreted it in terms of agent dispositions, or "metagame" rules. Sellars' position includes both the game and the metagame, and he underlines the causal notion of "exemplification" instead of the intentional notion of "abstraction" when explaining the relation between moves and the system to which they belong (Sellars, 1963). Here Sellars is alluding to a rich internal model (albeit agent-opaque), which indirectly covaries with the possible moves in the environment. His semantics will be further explicated in section 3.3 of this chapter, also with regards to conceptual role semantics and account of "translation".

model. In looking for this realm, Brandom advocates full on translatability (in his words "reducibility") from normative terms to non-normative terms, on grounds that "conceptual norms implicit in discursive practices" have objectivity independent of any agent's dispositions (Brandom, 1994, p. 63). And the nature of the objectivity is normative, particularly that there is no purely descriptive content, and all content is derivative from normativity-laden vocabulary such as asserting, committing, evaluating, etc., products of social interactions and practices (Brandom, 1994, p. 61, 2001, p. 607).

While the scope that Brandom derives intentionality from is limited (to language and therefore social realm), the resulting intentionality he infers is also "stronger", involving not only alethic (truth giving, factual) but also "other" modalities, properties, in other words, the general category of normativity proper, being subject to norms. The notion of norms covers a wide variety of intelligible (human or animal) conduct. This is a bit too strong, because norms can be cashed out in terms of rules, practices, conventions, or standards ("normal"), in different interpretations of philosophers, would further apply to ethical, legal, communicative… norms. Thus, one can extrapolate variations: ethical rules, ethical practices, ethical conventions or ethical standards, communicative rules, communicative practices, communicative conventions, or communicative standards, and so on. However, when we say normativity in general, we are making a general comparison between other types of regularities and patterns. Even though both laws of nature and norms involve regularities, we say abiding by the latter is a matter of "rule following", or the normative capacities of agents (as also highlighted above by Sellars). There is a tension in Brandom, on the one hand, generalizing over all sorts of norms over all domains of application is too general, and he limits himself to the social and linguistic realm, and communicative practices. On the other hand, he wants to

make the switch from mere alethic (factual, truth giving) modality (see section 3.2.2 of this chapter below) to normativity proper, by means of the device of inferential roles of concepts, but it is not clear how his project is generalizable, as the conceptual norms are highly sophisticated, language based forms of attitudes and it seems that more primitive forms of normativity are also needed.

In any case, Brandom himself also admits the need for an overarching account that merges both the psychological, as well as the sociological origin of meaning (Brandom, 2001, p. 607)[42], we do not need to limit the scope of investigation of the link between action and inference, to language. We can say, concerning our original problem, utilizing modal properties to ground intentionality is common to "meaning as use" accounts, but the scope of modalities they admit, as well as the mechanisms and emergent levels they fathom differ. Here I compared Millikan, Brandom, Sellars and Wittgenstein, to show the commonality between "meaning as use" accounts of semantics is their indirect and mediated solutions to the problem of intentionality, concerning the aboutness of representations. This mediation may take a more causal-functional or normative-pragmatic turn. They agree on their denial of a direct correspondence as underlying intentionality. They also utilize some sort of modality to enable the transition between more directly causal operations to developmental, agent relative, attitude laden "uses".

These are not the only philosophers engaged in the program of accounting for intentionality. In fact, Millikan is also part of another tradition, of naturalizing semantics

---

[42] "I am very interested, for instance, in what I see as Hegel's way of combining social and developmental (though not, for pardonable chronological reasons, Wittgensteinian and Darwinian) considerations in his account of the nature and content of conceptual norms" (Brandom, 2001, p. 607).

and intentionality, along with other philosophers such as Fodor and Dretske (see section 3.2.3 of this chapter below). Naturalizing in this context means, incorporating the study of normatively intentional phenomena, semantics, representation, and concept use within the framework, methods, and vocabulary of natural sciences as part of the overarching ideal of the unity of science. Before this generation of philosophers, the naturalizing project was limited to using non-intentional descriptive or causal concepts to account for intentional phenomena (content) for some time. The reason I started with Sellarsian philosophers is because, as Brandom (2001, p. 600) makes the historical point, Sellars' project, along with other developments in the history of philosophy such as Kripke's modal semantics, helped to legitimize the philosophical use of modal vocabulary to ground intentionality of meaning. However, we can say, while Kripke's project worked out to clarify and develop the modal landscape, saving modal vocabulary from its previous obscurity, it did not really address the core legitimacy issue raised by empiricists from Hume to Quine. The issue consisted of questioning the license in using vocabulary that relied on counterfactual conditions in explanations, such as disposition terms like "fragile" or "stable", habit, natural laws or lawlike occurrences, even the fundamental notion of necessity or necessary connections. Goodman (1983) phrases the classical "Humean" objection in his famous "New riddle of induction":

> The problem of the validity of judgments about future or unknown cases arises, as Hume pointed out, because such judgments are neither reports of experience nor logical consequences of it. Predictions, of course, pertain to what has not yet been observed. And they cannot be logically inferred from what has been observed; for what has happened imposes no logical restrictions on what will happen. (p. 59); [Past regularity that established a habit] at best pertains only to the source of predictions, not their legitimacy; that he sets forth the circumstances under which we make given predictions-and in this sense explains why we make them-but leaves untouched the question of our license for making them. To trace origins, runs the old complaint, is not to establish validity: the real

> question is not why a prediction is in fact made but how it can be justified. (p. 60)

On the other hand, Quine (1960/2013) in his classical *Word and Object*, is skeptical of the scientific use of dispositions as subjunctive conditionals (hypotheticals), as well as intentional idioms and propositional attitudes, what he calls "indirect quotation", in other words, all non-extensional idioms, because they cannot be paraphrased in explicit canonical notation:

> The farther a disposition is from those that can confidently be pinned on molecular structure or something comparably firm, the more our talk of it tends to depend on a vague factor of "*caeteris paribus*." This factor was what was elusive likewise in indirect quotation; and it was what obstructed general analysis of the implicitly dispositional constructions represented by 'Tabby eats mice' (§36). That the idioms involving it remain useful is due to those clues to the scope of "*caeteris paribus*" that are afforded by the context or other special circumstances of the particular utterance. This is why the paraphrasing of such idioms into a satisfactorily explicit canonical notation is often practicable occasion by occasion and altogether hopeless idiom by idiom.
> From recent paragraphs it becomes evident not only that the subjunctive conditional has no place in an austere canonical notation for science… our renunciation seems only in keeping with the trend of science itself: the favoring of definite mechanisms, assumed or discovered, over unqualified causality. (p. 206)

As stated before, Kripke's modal semantics helped the cause in terms of articulating modal operators by an extensional semantics, interpreted as first order quantifiers ranging over possible worlds. This articulation helped to address the issue of articulating the lawlike generalizations and subjunctive or counterfactual conditionals in precise, extensional, descriptive vocabulary. For instance, what does it mean that a relation is necessary? It means that the relation holds in every possible world. And so on. This solution somewhat satisfied the structural requirement of Quine, at the least more than Quine's own solution of structures of stimulus responses.

It's not the case, however, that possible world semantics or other modal logics achieved this feat by means of reducing all modal operators into non-modal ones. The notion of "possible worlds" is still modal, after all. The modal semantics did not address the issue of the validity of claims to necessity, the problem of induction, and therefore the philosophical problem of the validity of lawlike generalizations (a problematic that Kripke himself also pursued in his take on Wittgenstein's rule following). Brandom (2001) observes:

> It cannot be that, faced with the choice between giving up the use of suspect modal locutions and showing them to be respectable by reducing them to ordinary extensional nonmodal ones, we did the latter. The Kripke semantics is not even a candidate for providing such a reduction, because it owes its extensional character to the introduction of new primitive notions, possible worlds and accessibility relations (and in the case of quantified modal languages, further apparatus permitting re-identification of individuals and across worlds) that are themselves richly modal (and whose deployment turns out to require further metaphysically nontrivial commitments concerning about what is essential and what accidental). (p. 598)

What Brandom here draws out is indeed the real challenge for the naturalized semantics program, albeit by different means. Particularly identification and re-identification of individuals, ranging over and application of vocabulary to real objects, is a problematic we will see later.

However, the challenge of reducing modal vocabulary to so called descriptive, non-modal vocabulary was addressed by Sellars' inferential role semantics, by means of refuting the original requirement. After all - as Quine also admits (1960/2013, p. 205)[43] - apart from surface structure of the "-ble" suffix, "soluble" is not much different than

[43] In his solution, Quine resorts to the ability to drop disposition concepts for extensional "conditioned responses" as well as "systematic paraphrases" in terms of theoretical structures of "stimulus-synonyms" and "microstructure" (1960/2013, p. 205).

“red” in relation to being a “disposition term”, on the basis of counterfactual implications we relativize its meaning to. Thus, it begs the question whether there are any concepts, whose meanings are exempt from such meaning relations. Furthermore, explicating, caching out the exact relations that make up the webs of semantic and epistemic consequences, minimum (necessary) and maximum (sufficient) conditions, counterfactual consequences require using modal operators.

> For 'structural' appeals to properties such as valence and features such as molecular orbitals explain physical or chemical behavior only in virtue of the modally robust consequences they have, e.g., for what *would* happen if two molecules having a valence of +1 combine with one molecule having a valence of -2”. (Brandom, 2001, p. 600)

We cannot use concepts without implicit acceptance of their conditions of applicability. And the conditions of the applicability of concepts seem to require modal operations by the users that utilize them.

> One can never be in a position in which the use of ordinary nonmodal terms such as 'green', 'rigid', and 'mass', 'cat', and 'mat', is in principle legitimate and unproblematic, while the use of modal terms is in principle questionable and problematic. For we cannot make sense of what is expressed by nonmodal vocabulary apart from all consideration of counterfactually robust consequences and implications… modality is already implicitly intricated in the very meanings of nonmodal terms. (Brandom, 2001, p. 601)

Agreed, but what sort of modalities? And how far do modal operators could take us in our quest of understanding the link between semantics and representation, and habitual action?

3.2.2 Virtues and limits of alethic modality

We have said that many recent functional or pragmatic (or “interactive possibilities” in the interactivist case) programs of explaining intentionality of mental or linguistic

representations utilize modal, naturalistic terminology, to refer to mechanisms or procedures that allow causal counterfactual or subjunctive consequence relations. There may be other sorts of modal operators, but the kind utilized in such accounts is at least alethic or factual modality. *Alethic* means entailing truth, adhering to facts. The contrast is with other modalities, for instance something may be ethically necessary but not factually so. Ethical or legal necessities do not entail truth, while epistemic, metaphysical necessities do. A proposition which has epistemic or metaphysical necessity is necessarily true.

Returning to the core problem of intentionality in general, let us see the virtues and limitations of the philosophical tools of alethic modalities solve the problem of intentionality. The philosopher's term intentionality, as the key feature of representations, mental or non-mental content, has three special aspects[44] that we want to be able to account for, possibly by means of modal terminology:

---

[44] These three aspects are common to the semantic content of representations and concept use, indeed all psychological content. The notion also applies to semantics of non-mental (external) representation, such as pictures, words, maps, sounds, signs, yet the intentionality of these consumable, non-mental representations is derivative, based upon the intentional interpretations of human thinkers.

"Intentional" was originally used by Brentano to distinguish mental phenomena from physical phenomena, in order to come up with a scientific framework for psychology, via the special characteristic of every mental state having and being individuated by their intentional objects: "Every mental phenomenon is characterized by what the scholastics of the Middle Ages referred to as the intentional ... inexistence of the mental, and what we, although with not quite unambiguous expressions, would call relation to a content, direction upon an object ... or immanent objectivity". (Brentano, 2012) More clearly; "the existence of the intentional object *in* the state of mind itself…the object individuates the state of mind: different intentional object, different state of mind." (Crane, 2001, p. 29)

The origins of Brentano's concept of intentionality can be found in scholastic and early modern theories of ideas. The notion of intentionality is related to the act (thinking) - content (ideas) distinction of early modern period, in taking the mind's directedness at ideas as the paradigm model. In Descartes for instance, the ideas by virtue of having content within our mind have a being, perfection on their own right, which is their objective reality. This is not the thinking process; it is the thought. It is not the actual thing itself; it is the idea of the thing, it is its conceived essence in the mind. The ideas do not have to correspond to actual things to have objective reality, conceivability is a kind of reality. The objective reality pertains to the object presented in and through the operation of the mind. It is the notion of an imminent, in itself reality of ideas (as opposed to their transcendent reality).(Crane, 2001; Moran, 2014)

(1) Representing under an aspect, in a specific way, aboutness, representational contentfulness, intensionality, non-extensionality (for instance, “the morning star”, “the evening star”).

This feature was most clearly identified with Frege’s sense, as distinct from reference, or intensional contexts of substitution, as distinct from extensional contexts of substitution of co-referring expressions.

This feature is problematized in terms of obscurity and evasiveness. This was the characteristic of semantics that made Quine want to give up intensional states altogether, along with synonymy, necessity and analyticity[45].(Quine, 1960/2013; 1953) This is also the feature that eliminativists in cognitive science may be willing to dismiss first.

(2) Directedness, reference, pointing to or locating the intentional object within the relevant environment or context, extensionality (for instance “Venus” is the extensional reference, the object that the descriptive content in *one* is referring to).

The old problem of accounting for empty or vacuous reference relates to explaining this feature of mental phenomena. As Brandom remarks, vacuous reference is possible only possible with intentional relation, it is not possible in a non-intentional relation such as "kicking". "And one can want to buy a sloop, even though there is no particular sloop one wants to buy-even though one wants only, as Quine puts it, relief from slooplessness. But one can only sail or sleep on some particular sloop" (Brandom, 2001, p. 589)

---

[45] Quine partly justifies his elimination on Chisholm’s criticism of Carnap’s reduction.

(3) The normative force of intentional content, that content can be mistaken, concept use may be inappropriate, representations may misrepresent, there is a right way to follow rules, thus intentional meaning is subject to normative standards. The philosophical problems associated with this feature are rule following and accounting for misrepresentations.

Modal vocabulary makes possible to account for (1) and (3) more directly, and (2) in a more indirect and roundabout way.[46] Meaning as use accounts, that make use of alethically modal notions to describe representing relationship, such as pragmatics or functionality directly respond to feature (1), representing under an aspect. If my theory of meaning is proper function, my representations are at base about proper function, proper function is the base aspect. If my theory of meaning is inferential relations and meta-linguistic rules as dispositions that dispose me to obey those rules, my representations are about the interpersonal "oughts" associated with linguistic and meta-linguistic rules, dispositions or the conventional inferential relations are the base aspect (depending on the account). If my theory of meaning is anticipations of interaction possibilities, then this is the base aspect that my representations are about and are derivative from. The understanding of substitutivity of representations, their individuation, their development from more ambiguous to more precise, all follow whatever modal phenomena emerges because of the procedures the theories posit. The vocabularies of function, anticipation, disposition, or inferential rules, already involve relationships of counterfactual possibilities.

---

[46] (2) concerns decomposition, recognition as, identification and re-identification, and is taken up in the last section of this chapter: "properties versus concepts".

In the pragmatist-functionalist accounts, different kinds of alethic modalities are taken to be translatable into each other, depending on the type of mechanism and emergent function (modality) presupposed. For instance, in the Peircean pragmatist and interactivist account, an agent's anticipation or predictions of possibilities of action in an environment would constitute epistemic possibility (or necessity if there aren't other possibilities in the specific counterfactual situation). On the other hand, functional capabilities of the agent in an environment would constitute metaphysical possibilities (or necessities). Since both cases converge on probabilities of success (or error) in a particular interaction, pragmatic accounts of habit pose a scenario whereby epistemic and metaphysical possibilities are translatable to each other with contextual reservations. One can also derive nomic possibilities/necessities, as a separate alethic modality that pertain to possibility/necessity due to laws and lawful behavior of nature. An interaction consists not only agentive action, but also lawful interaction with the environment. As a result, we can say alethic modalities form a cohesive level of modal relations, united by validity, and different from other modalities when we are concerned with literal meaning.[47]

---

[47] Here I use "literal meaning" with reference to being evaluable in terms of truth value: being true or false. Evaluability of the meaning of representations and communication is not limited to alethic norms of truth and falsity. There can be other kinds of errors and grounds to reject. A case in point is different kinds of normative reasons to refuse. Here is an example from a class lecture by J. Fritsche, concerning the comparison of Habermas with Searle on speech acts. One can refuse to conduct the request: "please bring me a glass of water", on different normative grounds, for instance:

> on grounds of falsity, i.e., saying "there is no water left in the building". (This is an example of what I refer to as "literal meaning" above).
> On grounds of insincerity, i.e., saying "the requestor is trying to make me go out of the room".
> On grounds of moral wrongness, i.e., saying "it is wrong for a professor to use their status to ask students for errands."

Failures mentioned also apply to general action, gestures, or other meaningful human behavior. The example demonstrates that, the notion of correspondence relates more widely to the norms of action systems that define the structure of possible interactions. I am limiting the discussion to alethic or literal meaning here.

The most serious problem of accounting for the normativity of intentionality (evaluable nature) in terms of alethic modalities is that, strictly speaking, modalities concern types and not tokens. Because the solution of modal operators addresses the intentional, representational aspects of representations, which cannot be identified with the individual vehicle – tokens of representations.

Classically conceived, the type-token problem of encodings is this: if (the realization of) an individual mental representation in the head is a token, it cannot strictly be a representation, because it would have to be interpreted as a representation by an interpreter, the interpretative link would have to go on forever, and this ("little person") interpreter cannot be the brain, for instance. On the other hand, if an individual mental representation is a type (and not a token), how can it participate in causal mental processes, how can it be causally efficacious? The idea of an individual entity in the head, that has both the structural properties that allow participation in chemical-mechanical processes in the brain (a paradigmatic vehicle) and which at the same time has the semantic and intentional properties of contentfulness is inconsistent in many respects.

The theories we are considering must be able to account for cases of ambiguity, where a stimulus may represent more than one type of representation, and misrepresentation, where the type of stimulus and the representation don't match. For instance, I know that the chirping noise that I am hearing is either of a cricket or a cicada. In this case, is the actual token of my representation, of a cricket, or a cicada, or both?

Another example is line drawing, where a line drawing artist achieves to draw, stroke by stroke, impossibly complex scenes like a meadow full of flowers or a stadium

full of people, where it is impossible to show the exactness of the detail. When we zoom in to look closely, we won't be able to see shapes of people or flowers in the distance. We get the effect of the detail, that there are thousands of people in the stadium, without actually seeing the detail. The artist achieves this by drawing the closest objects in detail to stand-out, and they draw the rest not by shape or pattern, but by shadows cast and "negative space", "drawing the space between things". So the artist *establishes* what the details are by easy to see, recognizable, closeup silhouettes, with details that stand out quickly easily, and the rest becomes easy to attach meaning to by the observer. In this case, is my representation of the silhouettes that I perceive in the back of the perspective, of people, or of strokes that give the effect of the detail?

In both examples, in order to accommodate my representation of the phenomena, together with the vehicle – content distinction, I need to postulate that I represent whatever I perceive, via strictly the same vehicles, in my head. If it is an image, I must have a token of an image in my head, or if it is a sound, I must have a token of a literal sound in my head, or if it is a thought of a disjunctive sentence, I must have the same sentence written in my head. The truth is, thinking in terms of vehicles that carry contents (the paradigmatic encoding assumption) is very problematic.

Finally consider our previous example about the colleague and the elevator. Your colleague thinks that it is necessary to show his card and press the elevator button, to get to (any) floor. You, on the other hand, wonder if the communal floors would require the card, so you wonder if the rule works for all cases. In this case, is your tokening of this thought, in Turkish, or in English, in both, or in a separate mentalese language? If it is a separate language, do you also keep the Turkish translation in your head? And so on.

Isomorphism is very well, but it cannot be isomorphisms all the way down, there must be a more fundamental organization from which isomorphisms are derivable.

All these examples are experiential and interactive, therefore there is some causal correlation at work. The agents are relying on past "counterfactually robust correlations" (Brandom, 2001, 591) that has occurred in the past and that they perceive (and expect) to hold in the current circumstances. However what grants reliability, correctness conditions is not only the chains of actual token-token occurrences. In other words, environmental properties cannot be explicitly represented by associated tokens (If they were they would have been static "encodings"). On the other hand, for representations to be causally efficacious, some sort of causal covariance in token form is also needed to be incorporated. Since any causal encounter involves tokens of stimuli, tokens of objects interacted with, etc., however since the corresponding representations cannot simply be token copies. In such an account, there still is an agent required, applies the type of modality (necessity or possibility) to the actual token stimulus at hand. When one wants to equate the type of modal representing with the token of representation, one does not leave room for misrepresentation or error. Brandom puts the problem as follows:

> while the use of modal vocabulary offers semantic theorists a direct purchase on the non-extensionality of intentionality, the use of that vocabulary does not in the same way serve directly to make its normativity intelligible. Another approach is to do just that: start with some familiar kind of normativity, and try to exhibit the normativity of intentionality as a species of it. (Brandom, p. 592)

On the other hand, when one remains strictly on the "type" end of the type – token spectrum, (like Brandom does), one is stuck in an intentional realm, and one cannot account for the word-world connections, as well as the causal efficacy that must take place as an agent (learns to) physically interact with the environment and interacts

successfully. This was the original problem with which I started this thesis; how can we account for the gradual gain in precision of representations, as an agent gains skill in action? How do we link skilled action, practice, with learning and the change of representations? As the rules that we can apply get stricter to become more precise, the agent's underlying representations are expected to change. However, those representations at some point must be linked with tokens, and changes in tokens. Thus, we return to our original problematic: how can we account for the psycho-linguistic link between rules as types, applicable possibility and necessities, which are really abstractions that involve modal concepts, and the tokens of interactings, as well as the actual functional capacities, the procedural organizations of the agents, that allow the types to be realized as tokens in actual interactions?

This is a justified criticism, but it only applies to encodingist accounts. The solution to it was anticipated in Sellars' distinction between rules as "abstractions" versus cognitive maps or picturing as "exemplification". Encodingism assumes that representations are tokens, and therefore cannot solve the type-token problem. The encodingist program which was the original inspiration from semantics to the cognitive science program happen to assume extra conditions that cause the identification of representations (types) with symbolic tokens. This is not necessary, and is indeed an unnecessary condition on a program, which otherwise provides some developments over the semantic model.

One differentiating factor of interactivism is in the way it solves the problem of the derivation of alethic modalities (possibilities) from the logic of procedures. What is new with the interactivist take on modality is the use of implicit definition and the logic of procedures in giving the alethic possibilities, a concretely exemplified, causal

structure that allows counterfactual operations to run on. We will return to this issue in the last part of this section but let us take a detour on the progression from semantics to cognitive science program, to better position the issue within the progression of ideas, as well as to highlight the important novelty of the idea of implicit definition as it applies to concrete procedures.

3.2.3 The cognitive science program

The cognitive science program, in the way espoused by philosophers such as Fodor and Dretske, along with Millikan, changed the approach to respond to the problems of semantics, from conceptual analysis to "proof-of-concept" conceptual engineering. If one can provide a conceptual proof of how semantic phenomena can come about, one can explain what intentionality and the mental, is, is the idea. I am taking this detour into the origins of cognitive science to show the philosophical significance of the program for the rest of philosophy and theory of semantics.

To the extent that cognitive science program takes tokens of encodings as representations in the head, be they directly given environmental inputs, sense data, sentences in the head, biological vehicles, cognitive science does not show any progress over and beyond classical semantics in philosophy. Arguably the situation is even worse for cognitive science because it is not philosophically informed, falling into the age-old homunculus problem, the need for a homunculus (a "little man") which would be required to interpret the tokens as types of meaningful categories, objects, achieving the feat of interpreting the tokens semantically. Such a theory has nothing to add over and above the classical theories of reference, that stay at the level of abstraction and take it

as given that ‘chair’ refers to chairs,[48] without accounting for the theory of reference and the mechanism that realizes that theory.

That is not to say that no cognitive science theory can achieve that. Philosophers like Fodor tried to export the semantic explananda from philosophy like the directedness (referentiality) and aboutness (contentfulness) of intentional representations and tried to account for it in terms of some sort of a translation into a paraphrase in contentfully rich, but also syntactically clear and precise mental language of thought (Fodor, 1978, p. 246).

Language of thought theory was modeled after the practices of rigourous “paraphrasing” into formal languages in philosophy and (translation as) compiling into machine language in a computer. Fodor also ran through with the consequences of the idea of an assumed similarity between the so-called syntax of formal logic, and the syntax of computer (machine code), and the grammatical syntax of natural languages, which was supposed to map on to a parallel similarity between the relevant semantics, in terms of reference, content and truth. So, the semantics of formal logical sentences, the semantics of machine language, and the semantics of natural languages, should be similar enough to work in the same ways. It is clear to see that the model is not only operating on computationalist assumptions, that take tokens of codes and their encoding and decoding into other systems of code as involving intentionality, the model also includes imported assumptions with regards to the nature of translation, and the syntax-

---

[48] The problem is not that ‘chair’ refers to chairs, because it is (trivially) true that it does at the level of abstraction and the assumed relevant context of operation. The problem is a confusion between a theory of reference that is initially taken as a black box and an admitted ignorance, and later forgetting about the assumption of the black box and cancelling out the black box to take the relationship as directly given. For one thing, it is not justified. For another, the move is hidden behind engineering jargon that does not do the philosophical work.

semantics-pragmatics distinction from foundationalist analytic philosophy and linguistics. In this sense, Fodor is a conservative "Granny's boy" (Dennett, 1991), who however tries to keep the structural vehicles as syntactical tokens of representation; so that "it's syntax—all the way down!". Here we arrive at the other end of the spectrum between structure to normativity.

Apart from the type -token problem that we have mentioned before, the imported concepts are also problematic. A use-theorist would be quick to point out, for one, what Fodor takes to be translating can mean different things. What Fodor clearly means is a mechanical compiling, a one-to-one decoding of one stand-in token for another. But the nature of semantic translating, that a human agent engages in, when they are paraphrasing a proposition into another, even one that is as formally precise and well-defined as "Something is a lion" (in section 3.1), cannot be explained as simple compiling. If that had been the case, variables ranging over objects, or the theory of reference would have already been solved. Instead, the use-theorist would offer role-based semantics which is finding the role match within a system of representation in another system of representation as translating.

Instead of translating as compiling, the use-theoretic idea is related to the idea of translating as commensurability: that understanding is closely related with meaning. The idea can best be exemplified by Wittgenstein's famous remark "if a lion could speak, we could not understand him". Understanding a language involves a common frame of reference, that Wittgenstein calls "way of life", while other philosophers have characterized as conceptual frameworks, paradigms, or rule-matrices, which (natural or formal) languages, as systems of manipulation, embody. Fodor finds this idea some sort of reductive (empiricist) verificationalism, and rejects it as follows: "theory of how you

understand a sentence can do double duty as a theory of what the sentence means; to confuse compiling with semantic interpretation (1978, p. 286). It is difficult to tell how Fodor separates the two, without encodingist assumptions that overly rely on tokens of representational stand-ins.

On another note, the distinction between syntax, semantics and pragmatics cannot be as clear-cut, as semantics is inseparably bound with pragmatics, while the way to interpret syntax can only take place by means of habitualized interpretations patterns that are common most of the time, and therefore commonly assumed. We will return to this issue in chapter 4.

The point of the detour was to demonstrate the other end of the naturalizing semantics research program, where syntax as encoded tokens take precedence. The hope of integrating "[classical] formal semantics (type a), the account of the logical syntax of the vehicles of representation (type c) and the psychology of reference (type b) all fit together" (Fodor, 1978, p. 246) [49] was what gave impetus to the cognitive science

---

[49] Fodor (1978) distinguishes between three kinds of theories that model representational and communicative phenomena, somewhat problematically:
"(a) a 'theory of reference' (and, more generally, a classical semantic theory) which consists of a function from the expressions of a language onto the objects which interpret them"; Fodor also calls this "formal semantics", referring to model theoretic extensional semantics, which is taken for granted, in other words which is not the explanation but the explananda here;
"(b) a theory of the mechanism which realizes the semantics, viz. the kind of psycho-logical theory which answers the question: 'what about a given language, or about the way that a given organism uses it, makes one or another semantic interpretation the right one for that language'?'"; "when and only when reference occurs successfully"; Fodor means here truth conditions, which are indeed unbounded and not really needed to account for truth value according to this thesis, and;
"(c) theories that operate in…psychologized logical syntax; "internal" ("canonical", "mental") representation (hence all varieties of [procedural semantics], properly construed, belong to type (c), as does practically all of modern linguistics and cognitive psychology).", this is the encodingist (or computationalist) language of thought theory, which Fodor associates with the cognitive science program in general.

According to Fodor, cognitive science is engaged in "providing a kind of psychologized logical syntax" instead of a theory of semantics for a language; accordingly, a theory of semantics that shows how to interpret the syntax is the missing link that nobody has achieved to solve. (1978, p. 246) Fodor

program as a science of intentional phenomena. The program largely moved away from computationalism and symbol manipulations, despite remaining encoding assumptions even in contemporary accounts.

The question is, if we exclude explanations in terms of encodings, any ideas of structural tokens as representations, as well as any related notions of Shannon information and encoded translation, what remains of the cognitive science program? Are there remaining truths, true to the spirit of that program, that are worth keeping?

The idea of concrete operations, as procedures, allow modal possibilities to be realized on concrete systems, when we switch to the implicit model of representing, instead of explicit tokens of syntactical encodings. This move would preserve both the concrete exemplification of causal learning, as well as abstraction of rules and possibilities, in the original spirit and ideals of cognitive science. Neither mental processing per se, nor inference, nor memory have to be given up. Rather, these concepts should be cleansed of encodingist, explicit defining assumptions, which are unnecessarily limiting for the cognitive science program. One need not postulate: "a computationally localist realization of propositional attitudes" (Matthews, 2010) [50] in

---

implicitly claims that the problem is unsolvable, because he takes any attempts to solve it in terms of encodingism, and encodingism is inconsistent when we run through with the conclusions of its premises.

[50] Encodingist representationalist accounts are unnecessarily specific and particularist with regards to the local realization of propositional attitudes. Nevertheless, it is still the received view, according to Matthews (2010):
"Thus, to the question, 'How are propositional attitudes realized computationally?' the Received View proposes a single, general answer, namely, that the propositional content of the attitude is explicitly represented by some specifiable data structure, while the attitude itself is realized by some specifiable computational process defined over that data structure. Thus, for example, what makes it true that I believe that it is sunny outside today is that there is a data structure that represents this putative state of affairs, and this data structure plays the particular causal functional role in my computationally characterized mental processes that beliefs in fact play."

"[To give] a very strong, but empirically unsupported claim about the computational realization of propositional attitudes. Propositional attitudes are mapped into computational states and processes in a

order to give a viable account of inference and mental processing. The extra assumption pertains to explicitly, locally defining content as individual tokens, instead of allowing for causally real but implicit characteristics to emerge.

3.2.4 Procedures and possibilities

The novelty that interactivism brings to the intentionality debate is a new source of intentionality, that of interaction as a system of control, and procedures as if-then possibilities. The strength of this account is, interaction being very general and ubiquitous in nature, and which can take simplistic to very complex forms, is able to address intentionality from zero to highly complex levels of intentionality. Let us proceed with the derivation of possibility operators (as an alethic modalities) from logic of (intelligible) action.

We have seen that in-themselves non-intentional modal terms can somewhat account for the special intentional aspects of mental phenomena, particularly of representations. The three aspects were (1) intension, representing under a particular contentful aspect (2) directedness, pointing to a particular mental object (3) normativity, being evaluable in terms of correctness, on both (1) content and (2) directedness.

We said somewhat because different philosophers try to account for this problem in different ways, the non-intentional modal vocabulary that they resort to, and the associated semantic theories they admit differ, as well as their scope of application, and

---

very specific way: the attitudes themselves are mapped onto certain computational processes, the propositional contents of these attitudes onto certain data structures over which these processes are defined. More precisely, there is a homomorphism between propositional attitudes and computational states and processes that preserves, at the level of the computational algorithm, the type distinction drawn at the intentional level between attitudes and their contents." (Matthews, 2010)

the weak points of their accounts. Table 1 shows a summary comparison of modal vocabulary that the philosophers use to explain the intentional and semantic characteristics of representation.

Table 1. An Index of Modal Terminologies Used to Account for Intentional Features of Representations

	Modal vocabulary	Source of semantic correctness	Type of representing
Millikan	Proper function	Etiological, evolutionary history	Implicit
Brandom	Commitments, assertions, obligations… pragmatics	Users' participation in normative assessibility of concepts, implicit social agreement	Implicit
Sellars	Dispositions & inferential role of concepts	Learning rules (meta-game) and following rules (game)	Implicit
Fodor	Language[51] (of thought)	Symbols with causal and semantic aspects[52]	Explicit
Bickhard	Anticipations of interactive flow	Failure or success of interactive flow & selection for action	Implicit

In general, modal vocabulary, being alethic (evaluable in terms of truth/falsity) themselves, allows to articulate counterfactually robust correlations - causal aspects of the emergence of representations in the world - while accounting for three aspects of intentionality that are specific to mental phenomena. We have also seen that, for

[51] Fodor's take on semantics and language is not explicitly modal. He openly admits that he takes the semantic features of language granted, without having or attempting a semantic theory for them. Furthermore, in his arguments, he relies on the richness of natural languages, and how any mental language (or other processing) should be as rich as natural languages, while also being as clear and strict as formal languages. He relies on an ideal mental language that has both the reliability of computer code (or formal languages) as well as the semantic depth of natural languages.

[52] Arguably, this is not a source of normativity but derivative or attributed normativity.

counterfactually robust correlations not to be stuck on the type – token problem of mental representations, the normative aspect, source of normativity should be exemplified, instead of merely attributed as an abstract type of rule or an explanatory gloss. Meanwhile, exemplification can be functionally realized and causally efficacious, it does not have to be structural.

Arguably, procedures, the logic of if-then possibilities, meet these criteria best, at the sufficient level of generality needed. Basically, procedures have both the modal aspect of possibility, the earlier stages of procedures indicate the later stages as future possibilities, albeit fallibly. Procedures also have the causal aspect of reliability because they have the organization of series of indices, which are not in-themselves representational.

A procedure can be described as a control "flow",[53] basically a process with a determined start and end state, that keeps aligned or achieves a predetermined norm. In the interactivist model, an interacting agent, be it a simple organism or a human, must have the necessary competence, as the causally reliable process organization that allows the causal reliability, and they also must have the intentional aspect, implicit in the future possibilities of indicated further steps, to be able to interact with the environment. The indication can be sub-personal, does not have to be reflective in first level interactions. It is simply emergent in the indication from start to end; or preceding the later steps in an interaction. The control flow, on the other hand, is defined by

---

[53] The conventional term used is a "control structure", however the flow of control does not have to be structurally realized as the realization explicit state diagram, the notion of a structure is a metaphor in most cases. Control is a functional term, which is realized functionally, independent of physical and structural realization. Imagine an aquaduct system where the water level (the norm) is maintained by means of floating valves tied to open or close the emptying drains.

conformity to a norm, so it has indexical goal definition as an implicit end state (or range), functionally presupposed in the indexical[54] indication structure.

In talking about type-token problem of representations, we said that isomorphisms are very well, they are powerful stand-in relations, they are symmetrical. There is a symmetry in an isomorphism absent in a representation, if A is isomorphic to B, B is isomorphic to A, but if A represents B, B doesn't represent A.[55] Being symmetrical, isomorphic relations cannot be foundational for representation, there cannot be isomorphisms all the way down. There must be a level where another type of regularity and organization is operative, and from which isomorphisms are asymmetrically derivable. In the interactivist model, that organization is the logic of procedures whereby indications are implicitly realized via transitions and relations of indices. Indices (indexes) are tokens, exemplified, to be functional. In their exemplification as indices, in their causal operation, they are explicit tokens that function as pointers based on their token characteristics. From a "higher" emergent perspective of intervention,[56] they are indications for implicit differentiation, because a

---

[54] An indexical is merely a pointer, an index, that directs. An example would be a pointer in a computer code where getting to a state directs or points to the triggering of a separate function or assigning a new variable. Indexicals can be mechanically realized and are not themselves representational, do not have intentionality.

[55] Compare symmetry with transitivity: both concern 'alikeness', with regards to a comparison method. But transitivity concerns an order relation, being comparable on the same grounds, on the same *property*. The same relation holding when compared with a third object, so being comparable on the same property. Symmetry on the other hand, depends on the order of the comparison method. (Díez, 1997)
In the case above, representing is the comparison method, so my picture represents me, but I don't represent my picture. The metro stations are isomorphic to the metro map, and the metro map is isomorphic to the metro stations.

[56] Where does the difference between perspectives emerge? According to this pragmatist account, differences in perspective is a result of different possibilities of intervention. (A possibility may be a theoretical postulate and not actual. It is sufficient that it is conceivable.) A level is defined as an independent level of intervention or access. For instance, to treat depression, one can (attempt to) intervene with the chemical processes of a patient by medication, or one can intervene by changing the

web of indices may realize an indication, there may be nested webs of indices. A procedural flow may be emergent from (may be realized) by a web of indices, and the flow of transitions between them. From such a functional organization and implicit presuppositions emerges possibilities of interaction. When an index is pointed to, that index may indicate other webs of indices. The other webs of indices may function as a whole and make up a capacity and possibility to interact.

A control structure as a feedback loop is index based and therefore not representational. Any system that regulates according to goal criteria would fit this description, such as the Watt governor (Van Gelder, 1995), temperature regulation in the body, powered flight stabilization systems, etc. In fact, the notion of control is more general than feedback loops, they simply refer to modulation of one process to another, however feedback loops are what illustrate the importance of the general notion of control, apart from importance and the application of laws, as some machines cannot operate without feedback loops which are called "automated" control. An example in point:

> Many basic devices must be manufactured in such a way that their behavior can be modified by means of some external control. Generally, the same effect cannot be brought about (in practice and sometimes even in theory) by any intrinsic modification of the characteristics of the device. For example, transistor amplifiers introduce intolerable distortion in sound systems when used alone, but properly modified by a feedback control system they can achieve any desired degree of fidelity. Another example involves powered flight. Early pioneers failed, not because of their ignorance of the laws of aerodynamics but because they did not realize the need for control and were unaware of the basic principles of stabilizing an inherently unstable device by means of control. Jet aircraft cannot be operated without automatic control to aid the pilot, and control is equally critical for helicopters. (Kalman, 2022)

---

implicit presuppositions of the patient by therapy, or one can intervene by changing behavioral habits of the patient, one can even intervene on the physical level by administering electro-shock to the brain. Levels are defined by our methods of intervention, as intervention is our only direct access to reality.

As a result, the general concept of control is simply *external modification*, while feedback loops are a more illustrative possibility that the concept of control, when automated, allows. However, control, or automated control is not sufficient for representation or intentionality. One needs agenthood, in the sense that there is indication for the agent and selection by the agent, of action possibilities. This also states the nature of error or falsity conditions, for the agent, because it is the failure of anticipations.

Here we have talked about three levels of functional emergence (realized on whatever physical or chemical means) via procedures:

The zeroth level[57] is the functional emergence of control, by means of indexes and indicators, a simple mechanism of control which can also take the form of a more sophisticated feedback loop as automated control. At this level there is mere sequence but no indication for an agent, it is a blind but functional level.[58] It is a functional organization of control, not a structural or physical organization.[59]

The first level is the emergence of representation, where action possibilities for an agent in an environment are perceivable, by means of competence and goals. The

---

[57] I refer to this level as the "zeroth" because in the interactivist framework, knowing levels start with representation. At this level there is no representation yet.

[58] We can also call this "proper function", as Millikan does.

[59] "The concept of control is that any process that influences or determines the course of another process is said to control that second process. That second process may, in turn, serve to control still other processes. The organization of potential such control influences in a system is said to be the control structure in that system. Note that the concept of a control process is emergent from the concept of a physical process in the sense that any particular control structure organization can in principle be realized in an indefinite number of different possible physical systems; it is the functional organization of the control influences that counts, not their physical instantiation". (Bickhard, 1980)

underlying indexical organization flow enables this as a goal competence, but representation does not emerge until the competence provides an anticipation of possibility for acting for the agent. The possibility may also present possibility for action selection, or whether to act or not. Furthermore, the possibility also presents truth and falsity of anticipations, that may become falsified by a failure of the anticipated procedural flow. This is the level where animals interact, or humans interact out of habit or conditioning.

The second level is the emergence of reflective consciousness,[60] where the agent gains the ability to interact with their own control flow, and explicitly represent the possibilities of action, anticipations, failures, etc.

As a result, mental representations are indications of possibilities of interaction for agents, because representing emerges upon procedures. However, differentiating between cognitive processes (zeroth level) and representational level where we can talk about anticipations makes all the difference from other accounts of representation. Incidentally, this is what we talked about when we talked about the difference between contact (the causal flow) and content (representation) in Chapter 2.

The difference concerns an indication of the next step in a procedure and the next step in a procedure. The former one is from the perspective of the agent, and the second one is attributive, from an external or spectator point of view. What the agent knows is the former one. What the agent knows, strictly speaking is not the next step procedure, but the indication of possibilities of action. So, the indexical procedure is not transparent

---

[60] This is to note that first level and second level of representing are different. I will not get into the details of reflective consciousness and second level knowing in the Interactivist model for reasons of brevity.

to the agent. What is transparent is what the agent can do, can engage in, in the environment, provided by means not transparent to the agent, the indexical control structure that covaries with environmental redundancies that make interaction possible.

When we take a procedure (as a functional organization of control that belongs to an agent), to indicate if-then conditionals for the agent, the indications are by nature modal for the agent. That is, they are either possibly, or necessarily true, or false. Let us imagine a procedure as an algorithm, that involves webs and loops of if-then conditionals. Since a procedure is sequential, with a beginning and an end, and since the beginning (where one starts from within the web of if-then conditionals) determines probabilistically the following steps, indications of interaction possibilities that indicate further indications of further interaction possibilities define a course of counterfactual situations. Depending on the range of options, what is defined may even be a strict "nomic" or lawful necessity. Analogous to Kant's transcendental argument, an agent's possibility of interacting in an environment entails the agent's capability to interact as such, which in turn entails the agent's representing it. Because conditions of the possibility and thus capability of interaction in an environment for an agent also entails action indication and selection, which entails anticipation of it. This is the relation where epistemic possibility (/necessity), and metaphysical possibility (/necessity) are exemplified by the same phenomena of procedures. So, this is where the two (even three, considering lawful necessity) overlap.

First knowing level goal competence (norms of functionality, residing in effectiveness of functional control flow organization) and goal directedness (hierarchy of goals, normativity, being subject to certain ends) are supported by the definition of a control procedure. "The distinction between successful and unsuccessful interactions

must be made in terms of some sort of goal for the interaction." This is basic to the nature of a procedure flow that has a beginning state and an end state, and is not derivable from any declarative descriptions of the system. Noting that the control structure only becomes representation when the goal competence and goal directedness become indications and purpose for the agent, interactive models define knowledge to "consist of the ability to transform situations in the environment in accordance with goal criteria" (Bickhard, 1980, p. 11) or in short "task capability" (Bickhard, 1980, p. 5).

On the other hand, the functional organizational flow that supports robust counterfactual relations can be just thought of as indices and index structures, series of signposts that determine the procedural sequence. Putnam, in "The nature of mental states" (1980), defending functionalism with regards to mental states, differentiates such a functional organization of a control flow, from its physical-chemical realization, as well as the from the agent's knowledge of the underlying functional organization (as we have differentiated above). He explains functional organization via the example of the functional state of feeling pain as follows:

> (Of course, a Turing Machine is simply a special kind of Probabilistic Automaton, one with transition probabilities 0, 1.) I shall assume the notion of a Probabilistic Automaton has been generalized to allow for "sensory inputs" and "motor outputs"— that is, the Machine Table specifies, for every possible combination of a "state" and a complete set of "sensory inputs," an "instruction" which determines the probability of the next "state," and also the probabilities of the "motor outputs." (This replaces the idea of the Machine as printing on a tape.) I shall also assume that the physical realization of the sense organs responsible for the various inputs, and of the motor organs, is specified, but that the "states" and the "inputs" themselves are, as usual, specified only "implicitly"—i.e., by the set of transition probabilities given by the Machine Table. (p. 226)

This is akin to the interactivist idea: if we accept that "states" and "inputs" are specified only implicitly, by the set of transition probabilities. We can use the word "indication"

instead of “input”, their role is the same. Probability of the next state defined implicitly, gives “probability of motor output”. We might as well call “probability of motor output”, “interaction possibility”, as it is defined as the next state. The important point is that all of these are defined implicitly, by their role within the procedural if-then sequences or transition probabilities. Putnam (1980) calls them “descriptions of a system”:

> Since an empirically given system can simultaneously be a "physical realization" of many different Probabilistic Automata, I introduce the notion of a Description of a system. A Description of S where S is a system, is any true statement to the effect that S possesses distinct states S1, S2,...,Sn which are related to one another and to the motor outputs and sensory inputs by the transition probabilities given in such-and-such a Machine Table. (p. 226)

At a higher level of description, we may refer to these descriptions as functional states of the organism:

> The Machine Table mentioned in the Description will then be called the Functional Organization of S relative to that Description, and the S¡ such that S is in state S¡ at a given time will be called the Total State of S (at that time) relative to that Description. It should be noted that knowing the Total State of a system relative to a Description involves knowing a good deal about how the system is likely to "behave," given various combinations of sensory inputs, but does not involve knowing the physical realization of the S¡ as, e.g., physical-chemical states of the brain. The **S¡**, to repeat, are specified only *implicitly* by the Description—i.e., specified *only* by the set of transition probabilities given in the Machine Table. (p. 226)

We might say functionalism is a kind of proceduralism[61]. By demonstrating how a functional organization as a whole describe (implicitly define) a higher-level functional

[61] As mentioned before via the example of the carburetor, merely functional or procedural organization is not sufficient for normativity and representing to emerge just yet. Nevertheless, proceduralism is still a step towards programs instead of statements, an attempt to account for dynamism, and considering the possibilities admitted as ontologically real, instead of only considering actualities.

state, Putnam is also arguing against the decomposition critique against proceduralist theories. It is true that procedures may not always admit meaningful decomposition into sub-actions. In general, implicitly defined organizations do not admit decomposition, they define as wholes[62]. The part-whole relationship in procedural (or process based) functional emergence work differently, but as Putnam shows here, the causal structure and the differentiable identity of emergent functional phenomena is sufficiently well-defined, without needing to impose on it an artificial requirement to decompose or reduce as synonyms[63]. In order for an organism to be capable of a particular functional state (i.e., feeling pain), it would have to possess *at least* one description that determines implicitly the link between the capability to possess a functional state and the necessary functional organization that as a whole can be described as that functional state (Putnam, 1980, p. 227). We will get into the emergence of functional states and the related notion of causality, in the next section.

One caveat with Putnam's example as it pertains to pain is that his example collapses the index level and the first level, albeit the mechanics of a functional index

---

[62] Another version of this argument concerns decomposition into objects, events and agents: "[D]ecomposition is a critical problem for procedural theories which, in contrast to classical theories of objects and descriptions of objects (predicates), rely on what are in effect lists of operations or steps in a program. That is if we take our knowledge of the world to be based on objects, events, agents, and their interrelationships (a pictorial world image), then we may treat these entities as independent of each other and of ourselves more easily than If we must regard our knowledge of the world as a set, or a collection of sets, of expectations of what we will experience (what the total condition of our Index registers will be) when we interact with the world in each of all the ways we may act." (Younghouse, 1982, p. 56)

[63] The considerations that may lead one to assert that "water is h2O" as an empirical reduction, does not work for the superficially similar sentence "pain is a brain state". For one, "water is h2O" only works as a denotational reduction, in light of an associated empirical theory that justifies the shorthand. (Malt, 1994) The index structure, from which pain as functional state is emergent, is in effect an empirical position that is contrary to a shorthand that would allow us to assert "pain is a brain state". It would instead, support that "I shall, in short, argue that pain is not a brain state, in the sense of a physical-chemical state of the brain (or even the whole nervous system), but another kind of state entirely. I propose the hypothesis that pain, or the state of being in pain, is a functional state of a whole organism." (Putnam, 1980, p. 226)

organization and its higher-level description as a functional state recursively applies to both levels as implicitly defining. Just to be very clear, a control procedure may involve probabilistic features, may operate in ranges, etc. This does not by itself give truth or falsity, representation or normativity. The point made above about transition possibilities being alethic, capability of being true or false, only works with regards to an agent that takes the probable next step as an indication, whereby from the possibility of acting thus and so functionally emerges what we might call an "anticipation" or an "expectation", with regards to the first level interacting of the agent. Pure probabilities without action indication and selection are merely in the zeroth level, where there is no representation yet, but merely a causal reliable capability for control, which might as well be the navigation guidance system in a ship. So, there is no direct jump from a procedural organization with end state probabilities to representation. Which brings us precisely to a classical criticism of proceduralism, here articulated by Fodor (1978):

> … there is no way to distinguish between two systems that have the same procedures and states, but refer to different intentionalities and extensions, what is common between the two systems isn't interpreted meaning…Imagine two programs, one of which is a simulation of the Six Day War (so the referring expressions designate, e.g., tank divisions, jet planes, Egyptian soldiers, Moshe Dayan, etc. and the relational terms express bombing, surrounding, commanding, capturing, etc.): and the other of which simulates a chess game (so the referring expressions designate knights, bishops, pawns, etc. and the relational terms express threatening, checking, controlling, taking, etc.) (p. 232)

Fodor argues that it is in theory possible that, by accident, "the internal career of a machine running one program would be identical, step by step, to that of a machine running the other." Since all a machine language or an indexical procedure can talk about are "states and processes of the machine", the machines have no resources to refer to real references like "Moshe Dayan, cabbages, kings", can only refer to their names as expressions.

Fodor's argument is indeed correct, the index structure of states and processes will not give representation, at this level. This is not much different than attributing representation to causal signals. What Fodor here assumes to be representations are, are merely encodings. What could interpret the signals as semantic messages via a receiving system, whose representationality is built into what we take to be the conditions of the receiving system. In this example, representationality and the signal received are two different things, the signal is just moving the receiving system from one state to another state. All the other phenomena, references, intensions are emergent from the conditions of the receiving system.

As a result, the indexical level, information level, does not work as semantic information, it works as differentiation, detection, contact. The indexical structure is blind to what it represents, albeit being able to switch from state to state[64]. Not much different from a watt governor. Representation comes into the picture with anticipations.

It is clear to see that interacting involves procedures and possibilities of action, but it is not as intuitive for perception. In an interaction-based account, fundamentally perception also relies on procedural - interaction flows. Furthermore, the nature of these flows is transformational:

---

[64] That detection and tracking does not require representing is one of the major points of interactivism, as distinct from the received computationalist/ encodingist views. Fodor articulates the position as follows: "How is one to suppose that a mind can tell a C from a D unless it is already able to think about Cs and Ds? Likewise, how is one to suppose that a mind can infer from Cs being F to Cs being G unless it is already able to think about Cs (to say nothing of Fs and Gs)?... Tracking requires a way to represent the trackee. In fact, the point is quite general: epistemic capacities require ways of representing the intentional objects of epistemic attitudes. So epistemic capacities don't constitute concepts, but merely presuppose them." (Fodor, 2005)

The argument should be interpreted as a reductio as absurdum of classical (encodingist) representationalist accounts because based on the argument Fodor reaches the conclusion that acquisition of concepts is impossible. So, part of explaining the origin and acquisition of concepts is to account for how detecting and tracking is possible without representing.

> One genuine difference that seems to hold between Gibson's model and my own is that in mine apperceptions are based on interaction outcomes and the basic nature of those apperceptions does not depend on whether or not the interaction has accomplished primarily a detection or a transformation of the environment (or both). Gibson's perception, however, is a rather pure process of information pick-up with little or no effects on the environment. This fits our general intuitions about perception but leaves somewhat unclear such phenomena as walking around to get a better look at something or picking something up in order to examine it more closely and leaves such technologically aided detections as with infrascope or X-ray, seismograph or sonar, in a state of classificatory suspension. More fundamentally, a strictly nontransformational model of perception could not accommodate such phenomena as eye scans or head movements, phenomena whose importance Gibson rightly emphasizes. (Bickhard, 1980)

What makes procedural flows transformational and not merely static paths, or paths of transmission (of energy, of signals...)? What kind of a transformation is involved?

We might observe that the purposeful, goal-directed nature of procedures (and interactions) imply transformation, but it will not be sufficient. The transformational nature of perception is not due to instrumental reasons, because not all interactions and perceptions have to be instrumental. Transformation, instead, refers to the nature of representing, for which measure theory might provide a theoretical framework.

3.3 Representation as measure of intervention

We have talked about function and the importance of functional efficacy in explaining emergence of phenomena such as control, normativity, representation. An inadequacy with our philosophical understanding and current notions of function is that it seems vaguely verbal and descriptive. The notion of function does not seem to imply causal efficacy, connoting mere instrumental value, without real influence – in the way our collective understanding of efficient causality connotes influence. As a result, regardless of the many examples of emerging phenomena, regardless of the emergentist phrases in

natural language, function as an explanation, does not deliver the force of a satisfactory argument. It is as if functional explanation is causally inert, just a heuristic, an explanatory gloss, anthropocentric notion, even though it is not.[65]

I think the fault is not with the concept of function or the explanation it provides. The fault is with our concepts of cause and the wider conceptual framework. Without an adequate change to our concepts of cause, it is not possible to comprehend the causal influence of function, or emergence or emergent phenomena, properly. Without understanding these, we cannot understand representing, as our notions of causality also directly relate to the origin of representations and semantics.

The positions investigated so far has brought us to the following conclusion: the nature of representation is more a measure of intervention, of possibility space or indication for action, than a mirror of reality, (even though it can be utilized to function as a mirror or map of reality). Thus, understanding the nature of representation requires understanding what we mean by "intervention" and "measure". We will explicate a) interventionist account of causality, b) measure theory as it applies to representing.

---

[65] Block (1980, p. 171) differentiates between three kinds of functionalism in the literature:
(1) *functional analysis*: (Cummins, 1980, p. 185) looking at the overall function of a system by looking at its parts and how they work together to provide a functional explanation. This is a kind of psychological explanation. (2) *computation- representation functionalism*: psychological processes akin to workings of a computer program, composed of mechanical operations as primitive as the operations of a computer. Language of thought is systematically representing the world while inferences are computations. Computation – representation functionalism is a species of functional analysis. (Block, 1980; Egan, 2014)
(3) *metaphysical functionalism*: a theory of the nature of the mind, rather than a theory of psychological explanation, as in (1) and (2). According to this view "mental states are functional states" (Block, 1980 p. 172). (Block's own version of metaphysical functionalism is a functional state identity theory.)

According to Block (1980), what unites all kinds of functionalism is functional equivalence: "relations of functional equivalence for all versions of functionalism are relative to specification of inputs and outputs" (p. 273)

Metaphysical functionalism is the closest in kind to the type of functionalism we are interested in in this thesis, however Block's definition does not intrinsically involve normative aspects that would allow to account for normative features of representation.

Measure theory will ground the notion of *transformation* that is crucial for the interactivist account, as it pertains to not only purposive interaction, but also mere observation and perception. Interventionist theory of causality postulates that a semantic/explanatory level is simply a level of conceivable causal intervention. In other words, if we call domains of emergent phenomena levels (such as the psychological level, representative level, physical level, and so on), we will say that:

(1) a level is simply a scale of causal intervention,

(2) causal schemes are instrument relative, and

(3) instruments are function relative,

accomplishing the causal significance of function in terms of causal description and explanation.

When we derive representing from interacting, we can say that the problem of representation (and intentionality) boils down to a measurement problem. The measure theory concerns conversions (transformation) of units of comparison. The issue in question is, how one can "translate" what is perceived as warmth or heat into change of length, for instance. The so-called translation involves a transformation or a conversion. What is the nature of transforming what is experienced as "heaviness"[66] of objects into a uniform measure whereby we can talk about the characteristic without interacting with the objects? It would involve coming up with an instrument, inventing the "scale", whereby not only the idiosyncratic tool itself (which can be realized in many different forms), but also the function, the idea of "scale" is invented, whereby the characteristic

[66] "Heaviness" is also an abstraction of an invariance of interacting from all the ways heaviness is experienced in interacting, could be the quality of difficulty to carry, could be abundance of resources to use, could be destructive power when thrown, and so on.

of “heaviness” is isolated and independently reasoned over. Thus, measuring concerns converting possibilities of intervention into representative form, whereby we can reason directly over the representations. Any inference that we make over the measured units would counterfactually apply directly into interacting with the objects instead, so that reasoning over quantitative features replaces reasoning over interactions. This feature is called surrogative reasoning (Swoyer, 1991). Once a scale of intervention, instrument and unit of measure becomes familiar, an association can be formed so much so that the icon of a red thermometer becomes a symbol for heat. “Translation” as transformation is achieved by means of functionally corresponding the role within a domain of intervention with the role within another (logical) domain of intervention, and comparison. As a result, transformation (as conversion) is at the heart of representing, as it achieves the transition from the logic of procedures of intervention, a descriptive possibility space, into space of quantitative comparisons.

The strongest objection against this type of reasoning will be that such a pragmatist philosophy, that starts from interactions and interventions is anthropocentric, relativist, and therefore not epistemologically satisfactory. We will also attempt to address this problem of anthropocentrism of pragmatism, as that and the reality of function go together.

Let us start with what mean by intervention, which will lead to a better understanding of function.

3.3.1 Interventionist theory of causation and functional emergence

The core idea of the interventionist theory of causation is that our understanding is mediated via our modes of access (Bitbol, 2012). The idea of an “intervention” is that

controlled change leads to controlled effects, but this mode of control, mode of intervention, at the same time determines the predicates used to define the levels of access. The levels of access constitute reality. Bitbol (2012) defines intervention as follows:

> In ordinary circumstances, we are overwhelmed by the number of antecedents that are likely to have caused a certain event. Among all these antecedents, the one that we usually consider as the cause of this event is either the antecedent that appears to have changed recently, or, more convincingly, the one that we can get control of… Control, intervention, and action are central touchstones when causes are at stake. This elementary criterion for isolating causes then gives rise to a full-blown conception of causality, called the "interventionist" theory. According to this theory… configuration A is a cause of the distinct configuration B if: (i) whenever A has been actively set up by any means, B occurs (with probability p); (ii) whenever A has been actively removed, B does not occur (or occurs with probability $p' < p$)… [T]his definition of causality was retained by one of the founders of the modern science of nature, Galilei (1612, p. 425): "The cause is that which, when posited the effect follows, and when removed the effect is removed." (p. 239)

Interventionist theory of causation explain the main difference between mechanistic and (functional) emergentist explanations. Mechanistic explanation assumes independence from modes of access and involves a different criterion of objectivity. Emergentist explanation admits un-coincidence between levels of access.

An intervention is also a mechanism, so mechanistic explanation is not entirely wrong. It is just that it should not be absolutized, because mechanically described interventions are still, at base relative. Analogous to the way we say possibilities of action or affordances are relative to the agent and environment, an intervention is also already perspectival. The obvious issue that comes up in such an account is, objectivity. Objectivity is achieved through habit: patterns of controlled intervention constitute levels of description and controlled observation. The point is intervention is at the heart

of all explanation. Emergentist explanation is an attempt to be more explicit and openly perspectival about it.

As we have explained in Putnam's functionalism, the notion of implicit definition supports (explains the logical determining of) relativized access levels as patterns of organization. As a result, implicit definition is congenial to emergentist explanation. Explicit definition is achieved by converting implicitly defined modes of access into absolutized form.

3.3.2 Dualism and foundationalism

Bitbol (2012) uses the interventionist theory of causation to argue against mind- body dualism. His illustrative argument is comparable to Putnam's argument against reduction of functional states such as pain to brain states. According to Bitbol (2012), mind-body dualism is an outcome of a foundationalism with regards to the types of (metaphysical) entities presupposed. Identifying modes of cognitive access provide a unification of this dualism: the measure of free will is from the point of view of the actor, while the measure of mechanistic explanation is from the point of view of the spectator. Thus, when we say free will emerges from the physical body, what we are saying is two different modes of access/observation coincide in a (third common) indexical. I will argue that a similar reasoning can be applied in the case of intentionality and representation, and measure theory provides the method of the unifying indices.

Bitbol (2012) argues against using a dualist terminology between the emergent and the fundamental level of processes, because it ends up with a paradoxical concept of a so called "downward causation":

> Downward causation looks impossible as a concept, but is well established as a fact. A top-down flux of causation going from an emergent level of processes (say biological or mental) to the fundamental level of processes (say physical) that is supposed to underpin it, sounds like a paradox. Yet this paradox seems to be forced upon us when we want to make sense of many phenomena ranging from psychosomatics to the mutual interaction between organismic and cellular life; or when we want to express certain facts of experience in which a change in our mental state has an immediate impact on our physiological state. (p. 233)

Downward causation is defined as a causal relation where, a higher level (or emergent) process is said to causally influence a lower level (or fundamental) process, as in the typical example of mental processes influencing physical ones. A typical way the relation is articulated as a paradox is that in downward causation, a future state (a goal) is posited to influence a present state (Emmeche, Køppe & Stjernfelt, 2000).

According to Bitbol, the paradox is associated with two conceptual mistakes: first, assuming that the two levels are ontologically distinct, and second, by assuming that the two levels influence each other by means of efficient causality. We can see the assumption of efficient causality and the related paradoxes it leads to, for instance, in (Emmeche et al., 2000). The mistake lies in the ontological understanding of the levels, the physical level, the psychological level, the semantic level, and the inter-level causation between them:

> The metaphysical conception to be dispelled when downward causation is at stake is a dualist and foundationalist picture: a picture according to which downward causation only makes sense if there are at least two (ontologically) distinct levels of processes, with their own specific entities and properties able to influence each other by way of efficient causality. This is also a picture in which, conversely, if there exists only one fundamental level of being and process, then the very concept of inter-level causation is pointless. (Bitbol, 2012, p. 233)

The same conceptual issue applies to the emergence of mental representation. Computationalism is an attempt in the same conceptual framework of ontological levels and efficient causality: attempting to provide a solution to intentionality by means of

symbol manipulations, relying on the assumption of a symbol which is differently causally efficacious on the physical and semantic levels, which brings up the question of the source of causal efficacy.

Bitbol (2012) attempts to solve the paradox of downward causation by means of the interventionist account of causation:

> (i) Downward and upward causation are neither illusory nor inherently existent, but rather indexed by a certain level of intervention. Accordingly, downward and upward causation are not substantial concepts, but rather relational concepts throughout. (p. 234)

Similarly, the difference between implicit and explicit representing is not structural, but rather relational (in the sense of functional), indexed by a certain level of use. We will return to this point.

> Point (i) amounts to replacing ontological dualism with functional duality of domains of intervention. (p. 234)

Functional differences pertain to action, intervention, interaction, as defining the relations of terms. The ways of interacting with a view towards intervening, constitute the terms whereby we conceptualize and refer:

> (ii) The relevant relations do not take place between otherwise self-subsistent entities. They contribute to the very definition of their terms. These defining relations are primarily the relations between the agents and the targets of their actions; and secondarily the relations between various domains of acting and experimenting… As for point (ii), it dispels foundationalism from the outset. Instead of the usual scheme of intrinsically "basic" and "emergent" levels of being (or levels of organization), one is left with flexible levels of (experiential and experimental) access, levels of action, levels of analysis, and levels of theorization.

Here Bitbol (2012) is arguing against the terminology of "upward" and "downward" causation, as the analogy to higher and lower levels constitute the mistaken conceptual

framework. This kind of thinking imagines one level as foundationally substantial, ontologically made up of substances and properties, (or foundationally structural in the representation case), while the higher levels are made up "pure patterns" and configurations. Bitbol (2012) identifies this as a particular (foundationalist) framework of thought whereby "talk of upward or downward causation is at the same time irresistible and reduced to the status of a purely verbal machinery." (p. 326) Notice the parallel argument in Egan's (2014) criticism of computationalist (encodingist) accounts, where an explanatory gloss is disguised as grounds (foundations) of determinacy. Egan (2014) states "the mere existence of the causal relation (or of an appropriate homomorphism) would not itself justify the content ascription in the absence of the appropriate use." Egan diagnoses that it is often the case that the terminologies of representation (and other intentional variants) are not real attributions of content. The use of terms such as representation and content in such contexts should be taken to "play an expository role", an explanatory gloss to analogically explain what these systems are formally doing. See her commentary on the AI system Visua:

> …Visua is able to accomplish this, in part, because its states covary with changes in depth, or edges, in the immediate environment. The computational theory that characterizes Visua describes it in informal terms as *representing edges*. (Egan, 2014)

The implementational (algorithmic, physical) descriptions work to solve problems in counterfactual situations. It is left to the theory to explain whether the implementational content applies to the case in question by means of the intentional, explanatory gloss.

As a result, when philosophers really try to cash out the link between substantial and configurational levels, "either ontological dualism or ontological monism with epiphenomenalism" (Bitbol, 2012) is forced, which should be unacceptable for anyone

who takes the project of accounting for phenomena such as consciousness, intentionality, as well as quantum level phenomena, in a naturalistic way.

If we must keep to the terminology of "downward", we might attempt at a more precise formulation:

> We can then speak of "downward causation" in the following circumstances: (i) whenever an upper-level antecedent has been actively set up by means of a "coarse" instrument able to alter macroscopic patterns, certain elementary phenomena are observed by means of experimental devices designed for microscopic analysis; (ii) whenever this upper level antecedent has been actively removed by "coarse" instruments, those elementary phenomena are not observed. One can define "upward causation" likewise by just inverting the previous order of intervention and observation. (Bitbol, 2012, p.239)

The idea is simultaneity instead of subsequence. According to Bitbol (2012), an alteration of microscopic conditions by the macroscopic conditions has the concrete meaning of a disturbance introduced from outside, a disturbance which contributes to defining its own specific target through the scale and "coarseness" of the instrument one uses for triggering it. So "If one accepts that causing means altering a certain configuration rather than being [a level of] this configuration, one easily makes sense of such psychosomatic phenomena." The disturbance provides a common (third) way of referring to both microscopic and macroscopic change, change of configuration. Since the levels per se are not ontologically separate, the causal change is real. What the difference between the levels comes down to is an indexical difference of talking about the levels. We can contrast this to the idea of "weak emergence", where no real influence is supposed to take place, only the verbal, conceptual analysis differs, which is heuristic. What is articulated by the interventionist theory is not only verbal, weak emergence. Instead, through the achievement of the discovery of an intervention or a

measurement, a new way of referring to ontological change of configurations is discovered.

If the formulation still looks verbal and merely descriptive, we should verify if we still assess it in terms of "a substantialist construal of the basic elements" and efficient causes. Because this is what is undermined here. A substance framework is ultimately a foundationalist, base level framework: it assumes "individuable" entities as basic building blocks of other levels.[67] It is fundamentally ontologically asymmetrical. The point is to let go of the conceptual asymmetry between "thought, action, intervention, and higher levels of organization" and "the basic constituents" from which they are thought to be derivative from. This way of talking is not merely an account of the way we happen to know causes (epistemological), it is an ontological theory talking about what causes are. The concept of intervention is normally understood to be anthropocentric. The point here is that intervention is not more anthropocentric than other concepts, of so-called basic elements, of ontology. Our access and constitution of these spaces of explanation is constituted by intervention. Access is semantic and pragmatic. The claim is that the interventionist account is more realistic and truer to nature than so called scientific realism.

---

[67] Campbell and Bickhard (2011) make the point that it will not help the physicalist framework to resort to configurations of basic particles, because "a set of physical entities is not a physical entity, and physical properties can neither be affirmed nor denied of a set of physical entities." What makes particles, physicalist, basic entities, is that they are "elementary, they do not have a configuration". Their participation in a configuration, on the other hand, does not give particles causal power, because the (causal) properties that belong to the constituents of a set are different from the (causal) properties that belong to the set, the aggregate. Configurations are by nature holistic, "[s]ince they are holistic, configurations are properties of a set of constituents, …—not of those constituents themselves, nor of any aggregate of them. Ever since Bertrand Russell wrote his famous letter to Gottlob Frege in 1902, the logical perils of regarding a set and its members as being on the same level have been clear. It follows that configurations and their constituents cannot coherently be described in the same vocabulary." (p. 47)

> Without any ontological asymmetry, without any stubbornly substantialist construal of the so-called "basic" level, the dignity of the interventionist conception of causality is dramatically improved. On our construal, it becomes pointless to maintain a distinction between epistemological and ontological conceptions of causation, since nowhere can one point towards something like an ultimate inherent property that sets the standard of ontology. (Bitbol, 2012, p. 242)

There is a symmetry between epistemological and ontological levels when it comes to causation:

> At every level, the cognitive relations and the experimental interventions are constitutive of their object and of their target respectively. Therefore, it is not correct to claim that what is described as an intervention on a higher level of organization is in reality an intervention on its elementary parts. For these putative parts are no more inherently real than the upper levels of organization; they are just as much relative to a certain mode of investigation or action as these upper levels. At most, one can say that intervening on a high level of organization means acting on a domain of cognition that would be analyzed into parts if a microscopic mode of intervention were used instead. But this counterfactual statement should not be mixed up with an ontological statement about the elementary parts. (Bitbol, 2012, p. 242)

When one intervenes at the representational level, the intervention is relative to a cognitive mode of experimental access. We are considering counterfactuals, counterfactual references when we are trying to generalize to general causes. Therefore, causes are not "triggered" but constituted by interventions. "The physical (microscopic) cause is evoked only by counterfactual reference to a network of factors that could have been constituted by another (possibly incompatible) type of intervention." (Bitbol, 2012, p. 242)

We cannot draw out an in-principal distinction between "ontological" (absolute, efficient) causes and a "purely epistemological or anthropological" definition of causality, because the former is inseparably entangled with the latter, through experimental intervention procedures. And the way levels are drawn out is through the

consideration of counterfactual results of other types of interventions, constituting relational predicates. The so-called elementary level of micro-physics is a putative antecedent relative to a scale of intervention. “There is no cause that one could consider as a “primary quality” of matter; every cause arises as a “secondary quality” in so far as it is relative to a certain procedure of intervention.” (Bitbol, 2012, p. 247) The mistake lies in taking a relational predicate as an absolute one, to posit a foundation or achieve a reduction.

To summarize, levels of access are causal schemes (causal schemes are levels of access), constituted according to scales of intervention and observation. Constitution entails the three-fold claims that: (i) causal schemes are relative to “the method of active substitution of antecedent by means of various instruments adapted to various scales or levels”; (ii) accordingly, methods and instruments are the necessary presuppositions of causal schemes (iii) there is no fact of the matter as to which causal schemes are more real than others. (Bitbol, 2012, p. 248)

When we take the constitution of levels of access (causal schemes) by methods of controlled intervention and instruments seriously, the main issue to be solved becomes translatability, as well as causal closure. In fact, causal closure boils down to a translation problem since interventionist causality renounces absolute meta-observer perspective. What are initially separate scales of interventions and experimental access must be translated to be mutually consistent. This is further supported in the interventionist account that overcomes any dualism because there are no ontological differences between levels of access. Levels of access are not mutually exclusive and can be made mutually consistent through translation. There are no ontological

differences that would pause in principle reasons against translation. As a result, problem of causal closure boils down to a translation problem.

If every access is mediated through intervention, how different interventions can become reconciled is at issue. Measurement theory can provide a model to this problem if we render it as a measurement problem. The similar problem would be: how can one translate what is perceived as warmth or heat into change of length, into representative form? The representative form later becomes familiar and associates, so much so that the icon of a red thermometer becomes a symbol for heat. We will talk about this in the next section.

We can apply Bitbol's (2012) argument to representation and intentionality. When we talk about own representing activity, we are referring to a different mode of access (intervention/interaction/observation), then when we talk about truth and reliability conditions whereby these representings come about. The former is the actor mode of access, and the latter is the spectator. These are two perspectives that by definition cannot fully coincide, but they can be translated, partially, indirectly with reference to other (third) modes of access. When we talk about agent detectable error, again we are referring to a different mode of access whereby the agent detects that their representings had been wrong, mistaken, by means of a pragmatic, active failure to reach expected goals, which maps to a failure of truth value and not to truth conditions. We can try to derive truth conditions from truth values, but it is a different (translation) operation than direct converting of truth conditions into truth values.

Each level of access, determined by controllable conditions of intervention, has a conceptual unity within it. For instance, unity of an underlying procedural control structure, made up of indices may form a unity. Or the unity of apperceptive experience

may form another unity. Unifying two different levels is possible, but it requires conversion, we need a third, unifying level of access, that coincide in a common indexical.

The indexical level of access cannot merely be a way of speaking. The new common level must be, another level of access or intervention, on its own. That is what provides the realism as well as relevance to the existence of such a level. The levels are derivative from possibilities of controlled intervention, not merely attributed. What is at issue is determining "the boundaries of the constituting interactions between the basic elements". (Bitbol, 2012, p. 324; Emmeche et al., 2000) Here "determining" is considered a proper ontological influence that constitutes reality.

3.3.3 Measure theory and representing

Measure theory sheds light on the nature of transformation (or conversion) that occurs in the act of representing. This also converges with the interactivist model of communication, which is a transformation model.

The particular area of fundamental measure theory which talks about the foundations of conversion from experimental interaction to comparable scales is *metrization*. (Díez, 1997, p. 169) Metrization theory concerns mensurability conditions, facts obtained, or conditions that must hold empirically, for measurement practice to be possible. Fundamental measurement concerns measuring for the first time, when there is a conversion from qualitative empirical data to quantitative data is at stake. Derived

measurement concerns conversion of existing (often quantitative) measures to other measures.[68] (Díez, 1997, p. 169)

The significance of both fundamental and derived conversions for our case is that they concern making different levels of access mutually compatible, by positing a third level of access. It explicates how the third level of access that unites two different levels come about. Another significance of measure theory is that it demonstrates the tremendous utility of explicit external representations, in terms of the possibilities it allows for surrogate reasoning.

The idea of the application of measurement theory to mental representations is not new. Matthews, in his book, *The Measure of Mind: Propositional Attitudes and Their Attribution* (2010), develops an alternative measurement-theoretic account of propositional attitudes, whereby:

> the sentences by which we attribute propositional attitudes function semantically like the sentences by which we attribute a quantity of some physical magnitude (e.g., having a mass of 80 kilos). That is, in much the same way that we specify a quantity of some physical magnitude by means of its numerical representative on a measurement scale, we specify propositional attitude of a given type by means of its representative in a linguistically-defined measurement space. (Matthews, 2010)

Even though Matthews' characterization of the specification of (propositional) attitudes such as believing, asserting, denying…within a linguistically-defined measurement space is too abstract (not to mention, "attributive" and therefore closer to functional analysis in Cummins' sense), I think his core idea of relating measurement theory with

---

[68] "In derived measurement [mensurability] conditions are quantitative laws of nature", which in turn are studied by common empirical quantitative theories. (Díez, 1997, p. 169) This is why both measurement theory, and our exposition concerns fundamental measurement which isolates the relevant conversion operation.

representing is in the right direction. Matthews' account is too focused on (homo)morphic[69] relationships, and misses out on the pragmatic, conventional and interactionist aspects of representing. However, the core idea of measure theory whereby controlled empirical results are *converted* into consistent units (as measures) that Matthews analyzes keeps the interventionist link and illuminates the limits and relevance of transformation for representing. Particularly:

> Let A be a set of objects to which certain numbers are to be assigned representing the 'quantity' of a particular magnitude that they have. The facts related to the magnitude are expressed by certain empirical relations $R_{1,\ldots,} R_n$ (some of them can be operations) between the objects. Because the objects possess the magnitude 'in a more or less degree' some of these relations will be of (some type of) order. The domain and the relations make up an empirical system $\boldsymbol{E} = <A, R_{1,\ldots,} R_n>$ which expresses the essential nature of the property as a magnitude. Measurement assigns numbers to the objects, usually real numbers if the whole wealth of mathematics is to be applied. Empirical relations (and operations) $R_{1,\ldots,} R_n$ are represented by 'natural' numerical relations $S_{1,\ldots,} S_n$ which along with a set N of numbers . . . constitute a mathematical system $\boldsymbol{M} = <N, S_{1,\ldots,} S_n>$. The statement that numerical relations S_i represent empirical relations R_i means that $\boldsymbol{M}$ expresses with numbers what $\boldsymbol{E}$ expresses without them, i.e., that $\boldsymbol{E}$ is homomorphic to $\boldsymbol{M}$. (Díez 1997: 241–2)

Where this thesis diverges with Matthews, Diez and others is that they associate (and define) representing with homomorphisms. Therefore, their investigation into measure theory has the explicit aim of studying how homomorphisms are possible, that an empirical system must have to be able to measure it, and investigating the nature of the representation and uniqueness theorems that apply to measuring.

---

[69] Homomorphism requires/allows approximation because there is no strict one to one mapping, e.g., the conversion from analog to digital involves some loss of information, it allows for an arbitrarily small error, amount of loss of information allowed is based upon the unit of measurement chosen. Isomorphism is strict mapping between the elements of two sets, so isomorphism does not allow for approximation, and therefore is too strict for the representation relation. A lot of the times the purpose of representing is abstraction and idealization, leaving out unnecessary information. (Matthews, 2010, p. 138)

Our position as per the interactivist model is a little different. In our model, we already start the relationship of representing from the step of empirical interactions (operations) before homomorphisms and explicit representings come into place. Repetitive interaction (like practice or habit) conceptualizes a controlled intervention whereby an expectation of a regular outcome is born. It is only from the concept of an intervention as controllable outcome, that we can then derive scales as units. Measurement is a proper subset of representing in general. It provides an existence proof of how idiosyncratic content of interactive representation, can become agent-interaction independent and objective through systematic intervention.

The claim is that measure theory requires coming up with, and mapping between controlled levels of access, of intervention. One level of access is the empirical physical interaction with the scale – as a controlled, defined intervention. Another level of access is the intervention of using, manipulating numbers, number systems. How to translate the two? The solution is to come up with *a common unit*, the third level of access, which is the common indexical. The unit is an abstract structure, a model. There is some loss of information, but significantly relevant information is kept, whereby transformations coincide.

During the conversion, some information is omitted, and only information that pertains to empirically relevant relations is kept. For instance, concatenation is not appropriate to every kind of magnitude, such as measures of sensations. The use of different measurement scales such as is meant to capture differences in the allowable transformations from empirical to quantitative.

As a result, one of the core aspects of representing that it preserves information, particularly relevant, and potentially causally efficacious information, is accounted for. Transformation must be related to the preserving of information.

The chosen *units* in measure systems are exemplificatory models, based more on choice, than truth. The acceptance criteria of these models are truth as consistency in keeping the relevant information through homomorphisms, and convenience in application. Poincaré argues for the case that geometric axioms are implicit definitions, based on a comparison with measure systems, both of which are *models* based on "arbitrary" choice:

> In other words, the axioms of geometry (I do not speak of those of arithmetic) are merely disguised definitions. Then what are we to think of that question: Is the Euclidean geometry true? It has no meaning. As well ask whether the metric system is true and the old measures false. (Poincare, 1902/2012, p. 65)

We can make the same argument, in reverse, to mean that a measurement system with a chosen unit is an exemplificatory model, which is a definition in disguise, an implicit definition. The significance of this point is that exemplification comes before denotation.[70]

As has been argued in the interventionist theory of causation, measurement predicates are relational in their origin, while may look absolute in application. Matthews 2010) makes the same point:

> Having a certain weight and temperature are monadic properties of their possessor. The binary relations that figure in the logical forms of these sentences simply provide a way of specifying these properties, by relating their possessor

[70] A similar argument is made by Goodman (Goodman & Elgin, 1988) where he talks about exemplificational symbols: for instance a tailor's fabric samples exemplify a property, an aspect, without being denotational. Similarly, a work of art may exemplify mastery, or a mathematics proof may exemplify elegance. Here again, differentiating without denoting is illustrated.

> to certain abstract entities, viz., numbers on a scale, that are, in a sense to be made precise in the next chapter, the measurement-theoretic representatives of these properties. (p. 16)

In the sentences “Müge weights 60 kg’s” or “Müge’s temperature is 37.5 degrees Celsius”, one can say the predicates weight and temperature are monadic (one place) properties between a person and her property that belongs to her. That would miss out on the logical form of the predicates. The logical form of the predicates is relational, as they relate the possessor’s property to a scale of reference, which is another abstract structure. This was the same point made by Bitbol, with regards to causal predicates relate to the controlled methods and instruments of intervention, constituting causal levels of access. Indeed, a condition to properly grasp the meaning of the predicate is to understand its reference interval of possible application, counterfactual conditions of other possible values. Otherwise, mere numbers by themselves do not mean much.

Another point that becomes relevant here is surrogative reasoning. (Matthews, 2010; Swoyer, 1991) Measurement allows one to reason over possibilities of interaction and intervention, on an abstract level. It allows to

> reason indirectly about relations Ri among objects in the represented empirical domain X by reasoning directly about the surrogative relations Si among objects in the representational domain Y … furthermore that truth-preserving inferences defined over the surrogative relations Si will entail true conclusions about the relations Ri for which they are surrogates.” (Matthews, 2010)

I can reason that I can lift 20 kilos but not 40, and I would know for sure, without having to go through the actual endeavor. Surrogate reasoning allows more than simple ordering (comparison), in the most typical cases, it provides a system of reasoning where truth-preserving, seamless and complicated reasoning is made possible, that would not otherwise be possible. Strogatz (2019), for instance, talks about calculus as a system of

reasoning, where the use of the infinity principle sheds light on any continuous phenomenon such as shape, change, or motion. The symbolic vehicles of calculus do not have intrinsic powers themselves apart from how they are used or interpreted. The vehicles are significant but meaningless, unsaturated with meaning before someone reads them. We really are reasoning about the empirical things themselves, so the system of vehicles must approximate to another system (for instance the empirical features of reality) in order to be meaningful.

According to surrogate reasoning, meaning is a kind of coherence, provided by a kind of productive and informative, albeit "circular", reasoning. "Circularity" comes from exploiting equivalence relations, which are the preconditions of commensurability – mutual compatibility. That being said, the "circular reasoning" is not sterile, it is productive, because at base it utilizes implicitly defined models, and not explicitly defined stand-in equivalences. This may shed light on how implicit definition is congenial to emergentist explanation.

Models provide productive coherence, an informative circularity through implicit defining. Models are equivalent in some senses and not equivalent in others. They enable abstraction of features and idealization of approximations. They decrease computation effort to make certain implicit relations explicit - these relations may not be directly perceivable, but they can be inferred through operations.

3.4 Properties versus concepts

In the account we have been talking about so far, it should be noticed that representations are fundamentally intensional, they are emergent in the internal control flow of the agent. Control pertains to control of interaction. Representation emerges as

truth-laden, normative, epistemic only within the internal dynamics. As such, the externalist point of view can only be an attribution to an agent by another agent, a view from outside, or a god's eye point of view:

> It should be clear that such an interactive perspective also constitutes a strictly intensional perspective: if any conceptual categories or distinctions are made from any perspective other than the intensional interactive perspective of the knowing system, for example, if any reference to states of affairs in the environment is made in any way except through representations in the system of those states of affairs, then the implication is that either the system has direct knowing access to the world, in violation of the basic definitional assumptions of interactive knowing, or those categories and distinctions have not been, and perhaps cannot be, explicated within the interactive knowing framework but must instead be provided from outside the interactive model by the human knower who is considering the model. Such categories and distinctions would constitute, therefore, challenges to the conceptual adequacy of interactive knowing to all and, particularly, human, knowing. (Bickhard, 1980, p. 198)

This approach saves the account from veil of perception type problems; however, it raises some questions with regards to the nature of properties as distinct from the nature of concepts. Properties are supposed to be objective features of external objects, but now it seems that the interactivist account is reducing them to potentially agent relevant concepts. So, one problem seems to be, how the distinguish the one from the other, as well as how do we get from concepts to properties.

For in the received view, there are objects in the world, they have properties that belong to them, and we learn about the objects and the properties and develop concepts that pertain to both. We are supposedly aided by the extensions, the concrete objects that our representations range over. If our knowledge of the objects and the properties are constructions out of pragmatic intensional possibilities of interacting, how do we account for the objective neutrality of concepts as distinct from properties?

Furthermore, how do we account for extensions at all, or any extensional meaning that refer to sets of objects, sets of properties, in the world? How do we account for the extensions of our concepts, apart from their intension? Does intension make sense at all, without the notion of extension? Can this account be worth giving up on so many assumptions that we take for granted in philosophy, just to solve the mind – body problem?

3.4.1 The problem: Collapsing intensional concepts with properties

Overview of interrelated objections that surface to the account of concepts as implicitly defined are: (1) *reducing* properties to concepts, intensionally defining properties in terms of implicitly defined concepts, loss of direct reference and difficult to account for intension, sense[71] (2)descriptivism, *identifying* properties in terms of their descriptions only (3)This way of grounding properties gives at best descriptive differentiation, but cannot account for the phenomena of *re-identification* and *familiarity* (4) such coarse grained and general differentiations do not give the necessary level of detail to differentiate between *different types of* misrepresentation. Let's take them up in turn.

---

[71] Carnap's (1947/1988) distinction between intension and extension roughly corresponds to Frege's distinction between sense and reference: "Frege distinguishes for any name between its nominatum, i.e., the object named, and its sense, i.e., the way in which the object is given by it. We see from Frege's discussion that his concept of nominatum fulfils the principles of the name-relation stated earlier (24); thus his method of semantical analysis is a particular form of what we call the method of the name-relation. … Frege has made a very interesting distinction between the nominatum of an expression and its sense. … in the next section, compared with our distinction between extension and intension. It will be seen that in some respects there is a close similarity between the two kinds of distinctions; and it was, indeed, Frege's pair of concepts that first suggested to me the concepts of extension and intension as applied to designators in general. On the other hand, we shall find differences between the two conceptions, based chiefly upon the fact that Frege's conception is a particular form of what I have previously called the method of the name-relation." (p.118)

We say that implicit definitions only identify equivalence classes, they do not differentiate uniquely. There are instances when they can be explicated in terms of explicit or denotational objects (Carnap, 1947/1988)[72], however this is not necessary for interaction to succeed. Indeed, we have talked about this as a virtue of this account.

Suppose we have the properties P1, P2 and P3 in the environment. They may denote different properties, but to the extent that they satisfy the same implicit differentiations, they function as the same mental representation. Are not we in effect, collapsing the properties into organism relevant concepts, i.e., collapsing P1, P2 and P3 into C1? Putnam articulates it as follows:

> The effect of saying that the property P1 can be identical with the property P2 only if the terms P1 P2 are in some suitable sense "synonyms" is, to all intents and purposes, to collapse the two notions of "property" and "concept" into a single notion. The view that concepts (intensions) are the same as properties has been explicitly advocated by Carnap (e.g., in Meaning and Necessity). This seems an unfortunate view, since "temperature is mean molecular kinetic energy" appears to be a perfectly good example of a true statement of identity of properties, whereas "the concept of temperature is the same concept as the concept of mean molecular kinetic energy" is simply false. (Putnam, 1980)

Putnam is right to point out that the equivalences above do not sound right. Here he is arguing against equivocating functional properties with physical ones. We can briefly

---

[72] Carnap's understanding of reduction changed throughout the years. In Meaning and Necessity, he talked about implicit definitions as "contextual definitions", which were given in the reduction of intensions to extensions and the reduction of both intensions and extensions to entities. (Carnap, 1947/1988, p. 91) Here Carnap defines contextual definition as "a phrase like 'the class Blue' is not itself translated into a phrase in terms of properties; instead, a rule is given for transforming any sentence containing the phrase the class Blue' into a sentence referring only to properties."

Before in Testability and Meaning (1936), Carnap had developed a way of defining disposition terms such as "soluble" implicitly, based on Hilbert's method. He called such definitory sentences reduction sentences (R), which are, for him "open ended" partially determinate and more general than [explicit] definition sentences (D). We will revisit his treatment in the section below, however it is worth noting here the similarity of "disposition concepts" to implicit interactive anticipations for possibilities of action. Both express hypothetical possibilities, in other words concern logical modalities.

anticipate here that the copula “is”, do not refer to the same kind of equivalence in both cases. I take it that the copula “is”, does not denote a referential relation, but a role, within a representational system. And the roles of properties and concepts are different. We will return to this solution later.

On a similar note, one can object to say that derivation of extensions, identifying uniquely, accounting for reference is not intuitive, when we proceed from concepts to properties. On the one hand, we want to say that we want to be able to derive explicit representations, denotational objects from implicitly defined indications. On the other hand, it seems that this account does not provide any difference between extensional properties in the environment and intensional concepts, from the point of view of the interacting agent. Properties are supposed to have independent reality, regardless of our concepts, and we should be able to differentiate them as such. How do we account for this problem? A related criticism in this line of reasoning against Interactivist notion of fixing meaning of entities in terms of implicit definitions as equivalence classes instead of individuated particulars is “descriptivism”:

> Bickhard, because his representations are action orientated, is not a descriptivist in the way some philosophers are who try to cash out descriptive content in terms of sensory primitives. Instead he should be thought of as the pragmatist analogue of a descriptivist. The infant learns to recognize the red sphere by learning how it can be interacted with. But anything that can be interacted with in the same way will be counted as “that again”. So I take it that the interactivist account of object representations is an action based analogue of the satisfactional account of reference determination. We represent an object in terms of a web of action possibilities, and whatever we can interact with in the same way is what is represented. (Thorpe, 2021)

Here then, identification and re-identification of entities becomes a problem. Accounting for reference, fixing and tracking extension of particular objects should be more than a matter of giving determinate descriptions. Consider the condition of “agnosia”, where

individuals are able to perceive separate features or parts of objects but cannot recognize them as objects or identify the objects as the same. In this way, a person may lose their ability to interpret familiar objects or people. For instance, a person may look at their cat sitting on the windowsill, can see the cat's shape, color, and movement, but be unable to identify it as a familiar animal. Or cases where people lose their abilities to identify their loved ones from exact replicas or "clones".(Behrmann & Nishimura, 2010) The empirical examples point to a separate capacity to identify and re-identify other than differentiation via descriptive features.

> The function of substance concepts is not to classify things or to describe them, but to identify and re-identify them in the service of informed action…To possess such a concept, one must have a reliable capacity to re-identify the object in question, but this capacity of re-identification does not fix the reference of the concept. Rather the extension of the concept is determined by the function of the concept. (Matthews, 2006)

At this point, a glimpse of the solution becomes visible. Perhaps it is not descriptive features that directly express the identification conditions as a relation between representations and extra-representational reality. Instead, it is the functional roles, that determine what at some point was called significant concepts, that determine identification conditions. Maybe the problem is not so much the boiling down of extensions to intensions, but rather that taking 'intensions' as descriptions, instead of functional roles.

The fourth criticism was lack of differentiation between different types of failures: not all failures of interaction are due to misrepresentation. How can an agent differentiate between error due to misrepresentation vs. error due to bad technique? This criticism is especially pertinent, because according to interactivism, what differentiates one area of knowledge from another is error criteria.

To summarize, the family of problems to be addressed in the next section is as follows: (1) distinguishing between intensions of properties and intensions of concepts (2) derivation of extensions of concepts from intensions of concepts (3) identifying particular extensions and re-identifying them (4) Distinguishing between error conditions, i.e., error of misapplying representings vs interaction failure.

3.4.2 Partitions epistemology and constructed objects

Before we proceed with the inferentialist position, it is worth stating how much the interactivist position as a constructivism is accepting of some of the claims and consequences. According to Bickhard, representation is always in terms of (processual) differentiation. All differentiations at base are implicit differentiations, whereby appropriate satisfiers are said to interpret a sentence such that they make it true. So, the class of models that satisfy are unbounded: there is no a priori limit on the differentiations that we make, that pertain to the world:

> "The world is implicitly, but not denotationally defined, in that sense is opaque, but not unknowable, more and more can be known via experience, science and so on." (Bickhard, 2023)

Implicit defining also implicates that recognition of types is default in interactivism: types of interactions, types of processes. Representations of particulars are constructed in terms of criteria of further differentiation, such as of persistence, (re-)identification, etc. Unique identification is a further achievement that children learn after a certain stage. Unique identification has 3 preconditions: (1) the capacity of identification in the service of informed action, as distinct from; (2) identifying a particular, and (3) identifying or keeping track of individuals (re-identification or denotation).

One example that Bickhard gives is an anecdote from Piaget: his daughter mistakenly reidentifies a second snail on a path as the same as the snail seen before. Maybe children learn that (it is possible that) there can be many snails that look like each other, but people are unique. Another developmental evidence may be infants' attachment to the breast, or the care giver role, before they learn that mother is a uniquely identifiable object. In any case, the point of the example is to illustrate that identification of types, identification of particulars as particulars and reidentification are not readily "given" in the environment, children learn these, and people can make mistakes about these.

Of these, the development of identification of particular objects out of types of differentiations may seem puzzling. In the interactivist account, objects are not given, but they are also constructed (and learnt): object representations are emergent out of interactive representations. The epistemology of objects are constructive, based on closed web of interactions invariant under large class of interactions. Closed and invariant patterns of possible actions make up physical object representations.

One implication is that "relevance" is built into representations, as they are already construed out of webs of possible interactions, ie. the representation of the toy block is construed out of future potentialities of further interactions with it.

3.4.3 Inferentialism and functional roles

We have already hinted and anticipated the inferentialist – functional roles solution to the properties – concepts problem. An overview of the functionalist – inferentialist argument that will be presented is as follows:

(1) norms of accuracy are not sufficient to account for content of representations, also norms of obligation are needed

(2) functional role semantics is a case in point where a different role for the "is" copula in terms of functional role within a system as a translation can be sketched

(3) functional indices (as places in abstract structures) and not icons can enable denotational stand-ins to be able to account for extension.

(4) reidentification requires a representational system - to define functional roles within it

Before we complete the chapter with the inferentialist position, we must position the debate within the history of ideas to show its wider significance. What we have seen so far points to a domain of fundamental disagreement between preference over implicit versus explicit definitions, that lead to two different philosophies:

One, definitionalists[73], according to whom what gives meaning are explicit definitions and reference (ostensive definition) for primitive terms.

---

[73] The position of the definitionalists emerged as a reaction against defenders of (Carnap's (1936)) reduction sentences, more specifically disputing the status of so called reduction sentences as a legitimate way of introducing scientific predicates. Particularly, disposition concepts such as "soluble" or "magnetic" which cannot be straightforwardly defined in terms of observation sentences involving the material conditional, due to complexities with specifying time of events and referring to vacuous (empty) sets if the condition was not put to test. In Carnap's alternative, the reduction sentences could be solved if counted as axioms in the theory. Reduction sentences did not satisfy context independent syntactic eliminability and or non-creativity of definition statements. They would also allow incremental determination of meaning by means of introducing further and further reduction sentences for the same concepts:

"Thus, if we wish to introduce a new term into the language of science, we have to distinguish two cases. If the situation is such that we wish to fix the meaning of the new term once for all, then a definition is the appropriate form. On the other hand, if we wish to determine the meaning of the term at the present time for some cases only, leaving its further determination for other cases to decisions which we intend to make step by step, on the basis of empirical knowledge which we expect to obtain in the future, then the method of reduction is the appropriate one rather than that of a definition. A set of reduction pairs is a partial determination of meaning only and can therefore not be replaced by a definition. Only if we reach, by adding more and more reduction pairs, a stage in which all cases are determined, may we go over to the form of a definition." (Carnap, 1936, p.449) It should be noted that, disposition concepts are comparable to interactive indications and anticipations in terms of their modal characteristics.

Two, pragmatist and inferentialists, who utilize implicit definition to account for meaning.

Between these camps, the psycho-linguistic connection changes. The fundamental grounding and structure of meaning is different. The first camp takes the difference between sense and reference, object and attribute to be prior, more fundamental, and sometimes also as directly given. The latter camp takes it as derivative, constructed, constructable. Significantly the differences also imply further differences regarding interpretation assigned to predicates; the syntax – semantics distinction; and the dichotomy between analyticity and synthetic (between matters of logic and matters of fact); operationalism; as well as the proper subject of philosophy between context of justification and context of discovery. Attitude towards the project of naturalizing semantics is different. Role of reference is different. Account and justification of truth is different. Finally, the understanding of mental representation, in terms of encodings versus indications (as indices or pointers) is different.[74]

Returning to the dispute with regards to the ground of conceptual content:

> [O]n Sellars' view, conceptual content is not given by the rules of reference, as it is for the positivists/imperativists, (that is, the world-word connections), but rather by the word-word connections (rules of inference) it has with other words in the language (conceptual scheme). (Wilson, 2013, p. 533)

According to Wilson (1968) even if something like reduction sentences occurred, such expressions properly belonged to the psychological context of discovery and not the context of justification. And definitions should only concern the context of justification.

[74] The demarcation is not meant to group individual philosophers and reduce their intricate systems with sometimes incompatible terminology into one or the other camp, the above listed commitments may not always go together. Sketching up the two camps is justified only from the point of view of the development and consistency of ideas and progression of the general philosophical problem.

The different understanding of conceptual content maps directly onto the issue of primacy of explicit vs implicit definitions. We have said the operation of implicit defining concerns the relations between postulates, whereby the primitive terms are implicitly defined in terms of demarcating rules of applicability and satisfaction. This operation as guiding principle fits well with the inferentialist account of meaning whereby the rules of the application of utterances in different contexts are governed by whether the references in those contexts satisfy the norms of accuracy.

Exhibiting a uniformity and consistency of application is common to both rules and natural laws. However, only the former is normative, and not the latter, because of their imperative force. We understand rules to be normative, not because they involve a uniformity, a consistency in their application, (because laws also exhibit uniformity without being normative) but because of the normative force on the agent. The normative force here mentioned, “the ought” in Sellars’ terms, is not emotivist and motivational, but is rather conceptual and concerns conceptual meaning. Wilson’s comparison of Sellars’ “ought” with those of Ayer and Schlick is illustrative:

> For [Ayer and Schlick], conceptual content is given by the semantic rules of reference for the descriptive vocabulary. 'Ought' does not in this sense for these philosophers refer. Rather, its use on the one hand expresses (rather than describes) a mental state of feeling and on the other hand causes others to feel the same way. For Ayer and Schlick it expresses the (affective) mental state described by "I like that" and as it were urges you to like it too. This makes the view of these emotivists into a crude hedonism, for which they can rightly be criticized.
>
> They would have done better to endorse the more plausible view of Sellars (and Hume and R. M. Hare) that to say that something ought to be expresses the conative mental state of feeling that something is obligatory: we have a logic of imperatives rather than a logic of liking. One can then distinguish (as Ayer and Schlick cannot) between interest and duty. But to say this is not yet to allow that 'ought' lacks conceptual content: on the positivist/imperativist account it still lacks such content. On Sellars' view it does not…Since 'ought' has word-word connections, it will have conceptual meaning in addition to its motivational

> meaning which is primarily a matter of word-world connections… Sellars' view may be said to be one in which the rules of inference give meaning to the terms they govern by "implicitly defining" them. (Wilson, 2013, p. 533)

Thus, implicit defining is consistent with an inferentialist account where inferential relations between words that constitute and as well as constrain the conceptual. The structure of relations between words and sentences, make up the meaning of the words and sentences. Using the words and sentences brings with it a force of a rule, due to understanding them. The force is obligatory, merely based on the structure of the schema and engaging with the schema. (Sellars, 1963)

On another note, response to the distinction between concepts and properties would be Sellars' inferential semantics, inverse recursive differentiation. According to Sellars (1974, 1981), it is a mistake to take semantical concepts such as reference and aboutness are not relational, meaning is not a relation between representational systems (mental events) and extra-linguistic reality. He illustrates this point through the below sentences:

‘Rot’ (in German) means *red.*

can be properly paraphrased as:

‘Rot’s (in German) are ‘and’s.

In the latter sentence, the word is being mentioned, and not used. “Meaning is not a relation because 'means' is a specialized form of the copula, and, of course, the copula is not a relation word”. (Sellars, 1981, p. 329) “Means” is a special form of the copula (‘is’), indicating a role. Red is used in language English as Rot is used in German, not directly referring to redness. This essentially is not a direct relation between the concepts

red, rot, and is not mediated through redness. The role is acquired within a representational system, in this case a linguistic system (Sellars, 1974):

> Now if all this is correct, then our primary concepts pertaining to intentionality can be shown not to concern unique modes of relationships between mental events and reality, but rather to provide a technique for classifying mental events by reference to paradigms in our background language. (Sellars, 1981, p. 329)

A similar point applies to representational systems of non-human animals and other agents without language. He defines representational systems as a system of dispositions and propensities of an organism within which the organism locates itself and the environment: "to be a representational state, a state of an organism must be the manifestation of a system of dispositions and propensities by virtue of which the organism constructs maps of itself in its environment, and locates itself and its behavior on the map" (Sellars, 1981, p. 336) [75]

Sellars' conceptual role semantics is congenial to the theory theory of concepts. Concepts start their lives as stipulated theoretical entities, like stipulated scientific terms as implicit definitions based on lawlike statements in a scientific theory. They are posited in terms of the roles they are supposed to function within a theory. In a way they are transcendentally posited, as opposed to being posited according to what is

---

[75] Example of functional role semantics applied is artificial intelligence translation capability. Earlier models were supervised, given an English and a French version of the same sentence, whereby they inferred and gained the ability to chat with humans. Current versions like chat gpt are still supervised, however in a different way: given corpus of billions of sentences used in separate languages. They draw out a map (or a schema) of each word with associated words used within the same language. When they achieve a schema of language E and a different schema for language F, they are able to match words and sentences as "translations" based on the structural properties of the schemas, because schemas overlap sufficiently for the AI to be able to match individual sentences, without being given direct (explicit) translations of sentences.

Interestingly, the same method is being tried to decipher whale "language", where recordings of sperm whale bioacoustic data ("whale corpus of utterances") is given to the machine learning model, in order to attempt to translate to a human language. (Andreas et al., 2021)

observationally given. His story of the myth of Jones tells of a community that can posit the term 'thought' based on the model or analogy to overt speech, as 'inner speech'. That which is initially only posited stipulatively can then become observable, and this indeed is a condition for being able to observe, as concepts are abilities to attribute.

We must notice that the inferentialist position, as well as the interactivist model are congenial to the (ante-rem) structuralist position of Shapiro, in terms of defining objects according to their "location" in an abstract structure. Shapiro (1997) explains that we do not arrive at abstract structures (defined implicitly), through simple abstraction where one apprehends "small structures like letters, numerals, short strings" (p. 114), but instead:

> The important point is that we must leave the simple property/object dichotomy and think in terms of places in a pattern or structure. What the various "E"s have in common is that they all have the same role in an alphabet and in various strings. That is, our child has learned to recognize an alphabet structure and "E" as a place in it—the fifth place. (p. 114)

This already hints at identification and reidentification within a system. Shapiro (1997) demonstrates what identification can be realized in a system where there is a base language and sub-language. He calls this "Kraut sublanguage procedure":

> Our next step is to formulate a sublanguage of the base language for which the equivalence is a *congruence*. Two conditions are necessary: if F(x) is a predicate of the sublanguage and p, q are in the same equivalence class (i.e., the equivalence relation holds between them), then F(p) iff F(q). That is, members in the same equivalence class cannot be distinguished in the sublanguage. If two items are equivalent, they are indiscernible. Second, if two items m, n are not equivalent, then there should be a predicate Y(x) of the sublanguage that holds of one and fails of the other: Y(m) but not Y(n). (p. 123)

This is all within the same sublanguage. The sublanguage is impoverished with respect to the base language. When we move to the base language, what was indiscernible in the

sublanguage can become discernible through new predicates. “A structure” is made up of both the language and the sublanguages, where equivalence classes and their relations can be formulized. Accordingly, Shapiro suggests that the equivalence relation is the identity relation within a sublanguage, and objects become places in a structure. Another example from Shapiro (1997):

> A standard technique in algebra and number theory can be understood in these terms. Imagine a mathematician who decides to speak an impoverished language that cannot distinguish two integers if their difference is divisible by 7 (or, equivalently, if the numbers produce the same remainder when divided by 7). On her behalf, we make the indicated identifications: 2 is identified with 9, 16, -5, and so forth, whereas 3 is identified with 10. We interpret our mathematician as saying that 5 + 4 = 2. Of course, 5 + 4 = 9 as well because, in her system, 2 and 9 are indiscernible and thus identical. Under this interpretation, “identity” is what we call “congruence modulo 7.” The locution “2 = 9” can be rendered as “2 / 9 (mod 7)” in the background language. (p. 121)

From these examples, we can see how identity relations can be developed within a structural framework, through further discernment. Identity is kept so long as there is an equivalence in terms of the role within a structure.

CHAPTER 4

EXPLICIT VERSUS IMPLICIT REPRESENTATIONS

4.1 Embodied representing vs symbols that fail to represent

If you are not convinced until now that the act of representing at its core is not homomorphism, explicit definitions, or any stand-ins for that matter, (and instead are indications of action possibilities), I hope that you will be with the next examples that demonstrate that neither homomorphism, nor structure is sufficient or necessary for representing.

The following examples are from Kirsh (2006, 2009, 2011), that investigate the link between interpretability and the structural features of representations. The distinction has different uses in the psychology and philosophy literature, from "implicit knowing" that is meant to cover "tacit" or "unconscious" or "subconscious" knowledge (as in Chomsky's innate grammar theory), to "making it explicit" in Brandom's philosophy, where one uncovers inferential roles of concepts and the commitments agents are engaging in when using them, just to name a few. The purpose of this section is not to go through these different uses, to unite, explain or pick among them. The purpose is rather, an effort at understanding the true nature of representing and interpreting representations, to the extent that the distinction between dynamic implicit defining as equivalence classes and explicit defining as homomorphic encodings might be informative to understanding agentive representations as agents use and produce them, in learning, in creativity, in problem solving.

The issue of encodingism and how it is processed apart, let us keep with the literature and commonsensical usage, and use for the time being the terms "implicit" to

mean tacit, hidden, and implied, while "explicit" to mean openly present and readily graspable, to understand the functional nature of an agent's relationship to representations.

Kirsh makes the claim that the "implicitness" or the "explicitness" of representations is not in their linguistic or syntactical properties, but rather, it is in the amount of computing it takes for a consumer of the representations to interpret them. For instance, 'fifth root of 3125' is a less explicit [representation] of 5, than the numeral '5'.[76] "Hence the information is not *on the surface* in '3125' but is on '5'. It is implicit, but less so than $\sqrt[17]{762{,}939{,}453{,}125}$." (Kirsh, 2009) Kirsh is correct to reject the verbal understanding of representations, but his position should be even stronger: not only is the distinction not based on language and linguistic syntax, but the difference is also not in structure either. The difference is functional. In fact, Kirsh himself has further examples to illustrate this point. For instance, please read the below sentence, to assess whether it is implicit or explicit: (Kirsh, 2010)

POLICE POLICE POLICE POLICE POLICE

Note that it is grammatically correct and therefore structurally homomorphic to a proposition. So, the problem is not with structure, the problem is that it is not familiar or expectable for most people. The sentence is symbolically "explicit", but still requires deciphering into "Police who police other police who police other police". Still requires

[76] The original quote is "'fifth root of 3125' is a less explicit encoding of 5, than the numeral '5'." (Kirsh, 2006) Kirsh associates "computing" with the operation of encoding and decoding, and in cases of symbolic processing this may be partly operative as we will explicate below, however is mistaken as a theory of origin. Regardless of the encoding terminology, Kirsh's intuition with regards to the effort it takes to cognitively process, as well as the functional nature of the implicit – explicit distinction is consistent, even without accepting a strict encoding/decoding (isomorphic) framework of representation.

interpretation as to the goal of the sentence and how to parse it. The example shows that being syntactical is not sufficient for being explicitly graspable representation.

The goal of these examples is to demonstrate the importance of familiarity, as once one knows how to interpret, the next encounter will be significantly different. After one becomes too familiar with a complex phenomenon, perhaps memorizing some patterns, one can even interact directly with it. In line with this:

> Similarly, a creature with a highly parallel computational system, such as human vision or motor control, may be able to process complex structures rapidly when they are visually or motor encoded, but more slowly when linguistically encoded. So content shown visually might be explicit while being more implicit when given linguistically. It also gives a place for learning, since highly practised agents can immediately grasp contents, such as wine tastes, musical structures, concepts and so forth, that would be difficult for the unpractised. They have automatized or parallelized them. (Kirsh, 2009, p. 401)

The examples bring up the possibility of "functional encodings", which is a contradiction in terms. If we accept the arguments from previous chapters that the nature of an agent's representations cannot be structural, internal stand-ins that operate as simultaneously syntactical and semantical codes, that such kinds of objects cannot participate in the internal processes of an agent without a homunculus, then how do we account for external representations, real codes, that human cognizers are able to produce and consume? Are not these paradigmatic external stand-ins, internal stand-ins at the time they are produced, and before they are expressed? Even though being an encoding is an idealization, there are indeed "encoding-like" mental stand-ins, as well as epistemic, external stand-ins. To be able to properly give an account of their nature and relation to the fundamental kind of mental representing implicitly by means of , we need to clarify a few points. First has to do with "embodied" representations. The second has to do with the reversal of the terms "implicit" and "explicit", as in our framework, it

is the organization of interactive flow that allows functional indications to be directly interpretable and thus directly participate in the interactive organization of the agent. The interpretation is made possible by means of indicating implicitly. So, what is directly processed and more readily given is the internal organization of an agent which is implicitly determined. Whereby in common usage, and in Kirsh's writings, like in many others, what they term "explicit" is what is readily graspable[77] by an agent. Matched with an encodingist or homomorphic understanding of the nature of representations, the order of explanation is reversed. The problem of course, is how a homomorphism can be given. As in many other problems in the history of philosophy, the confusion comes to be what we take to be readily "given" and what we take to be "mediated".

To recap what we talked about in chapter 2, interactive notion of representation involves indexical indications that directly partake in organizations of action capabilities of agents. The agent's expectations of interaction outcomes in the form of procedural flows realize the normative aspects of representations. Agents can and do "define explicitly", but it is functionally not equivalent to implicit differentiating. An explicit definition is a specific stand-in, though this is an idealization. In fact, there are no context free, independent explicit definitions, there are no real encodings, but we approximate to them. At base these utilize implicit differentiations.

---

[77] The words Kirsh uses are "immediately graspable", "representable" and "expressible", "articulable" – to mean that can be thought, can be spoken rather more easily, in line with the commonsensical meaning of being open and obvious. Kirsh assumes the received encodingist view of representation (and knowledge) whereby representations have to be encoded as codes and decoded to be made explicit (therefore involving some sort of homomorphism).

An agent can also shift between internal stand-ins and implicit differentiatings via practice because the difference is in the way the agent is able to use, consume, knowingly interact with the representations, and not in the structural characteristics of the representations themselves. Since the difference is functional, the same representation, initially an explicit representation like a syntactical symbol, can become an implicit indication, one can learn to read the representation directly (as in Brandom's example of chicken sexing) or one can take an indication and try to make it explicit, to work and reflect on it, using it as a model of the original, direct indication. We will talk more on the function of internal stand-ins, but we need more examples first.

Let us use the less confusing terminology of direct and indirect, and continue with embodied representations, such as expressive dance moves. We agree with Kirsh that the degree of directness and indirectness is in the way receivers are accustomed to interacting with physical or abstract objects such as letters, numbers, artefacts and even their own limbs, their body parts. Another one of his illuminating examples is the phenomena of "marking" in dance:

"In dance, there is a practice called "marking". When dancers mark, they execute a dance phrase[78] in a simplified, schematic, or abstracted form." A marking is a kind of intended gesture by a dancer to study the real move by illustration. This is a significant example because it is both (1) embodied (2) representational. I will argue that a dancer's mark is an external representation of an implicit predicating. To understand the nature of this "embodied" representation we need to understand its use and purpose.

---

[78] A dance phrase is "a brief sequence of related movements that have a sense of continuity and artistic or rhythmic completion" in a dance performance.

> Based on our interviews with professional dancers in the classical, modern, and contemporary traditions, it is fair to assume that most dancers mark in the normal course of rehearsal and practice. When marking, dancers use their body-in-motion to represent some aspect of the full-out phrase they are thinking about. Their stated reason for marking is that it saves energy, avoids strenuous movement such as jumps, and sometimes it facilitates review of specific aspects of a phrase, such as tempo, movement sequence, or intention, all without the mental and physical complexity involved in creating a phrase full-out. It facilitates real-time reflection.

In Figure 5, an example of marking in Irish river dance is given. The movement of the dancer on the left is represented by the same move being marked, on the right. One hand is used to mark the movement of the steps while the other hand marks the floor. (Kirsh, 2011)

Figure 5. Marking in dance (Kirsh, 2011)

A significant facet of this example is that dancers not only use marking to talk about moves with each other, they also use it by themselves, to rehearse the moves, to think with their body. According to Kirsh (2011), marking for self is "the practice of creating a simplified version of a process – a personal model to work and think with":

> "... sometimes even when you know something it is good to mark it through - to rework it in your brain, to make another idea of what the movement is. (Dancer M in interview)"

The aim of the dancer while preparing for the performance is to find what her exact movements in the performance should be. Even though the general features of the dance

phrase are determined, how it will be expressed by the artist is a creative question. So in practice, the dancer is expressly looking for new ways to interact via and with her own body, and the expressive capabilities her body provides to fulfill the artistic intent of her part. How she can already execute the part, her current possibilities of movement, are her existing mental representations for the part. She is working to develop to change her existing representations, by means of variation and selection, by engaging in marking. Because in marking, she is working with "copies", "stand-ins", of indications, implicitly defined possibilities of what sequences are available to her at her current level of capacity. In other words, by marking, the artist is representing possibilities of what her current capabilities can develop into. For that reason, it is as if she is working with toy models. Even though these "models", these representations are embodied, and she is "thinking with her body". This is thinking because she is interacting with her own interactive flow, very similar to second (reflective level) of representing. Here she is utilizing the explicit stand-in process, to work out different alternatives, different from the interaction flow that she already possesses.

Based on the position developed in the chapter on representation as a measure of intervention, true creativity is coming up with new modes of access, new modes of interaction and transformation. For instance, what the dancers are looking for is new modes access, when looking for a new expression. A new form of access provides the new form of expression.

There is no such thing as objectively explicit, being explicit or implicit is not a structural characteristic of the object of interpretation, it has to do with the previous knowledge and habits of the agent making the decoding. It is a functional difference for

the agent doing the interpretation. One can interpret as an indication, directly, or one can interpret as a stand-in for an indication.

If there is no ultimate indication, one cannot interpret an explicit representation as a representation at all, for instance an ant cannot interpret the picture of a tree as a picture of a tree but only as a walkable surface.

4.2 Learning, framing and heuristics

Often, genuine learning is a matter of rational construal: it requires reframing or insight rather than deductively following deductive inference or systematic consideration. A heuristic may help as a shortcut, but where from?

Most of life's problems are ill-defined: If the agent does not know, either one or all of what constitutes the initial state, the goal state, the allowable operations they can make, the constraints that bind them, or the problem does not have a unique solution, such as; finding a perfect mate, writing a good paper, performing well a known dance sequence: "the evolution of Odile over the years". One needs to formulate first to solve them.

Heuristics may help with framing ill-defined problems. Heuristics are highly likely rules/procedures for success and relevance with no guarantee of either relevance or success, while algorithms are systematic with a certainty of relevance and success (Bickhard, 1980, p. 109, 213, 109)

The question is can they both be subject to intractabilities? Algorithms necessarily require frames of reference while heuristics work with ill-defined problems, too.

Relevance is defined for a goal-directed system when it functions as a problem solver. A problem-solving system by definition has an initial state A (situation), goal state B, and transformations (interactions) required to change A to B. Heuristics can act as preliminary models.

CHAPTER 5

FRAME PROBLEMS, UNITY, SYSTEMATICITY, PRODUCTIVITY

Let us apply the framework to some philosophical puzzles. The point of this section is to show that framing is prior to deduction.

5.1 Three puzzles

Three constructed logical puzzles, work in similar ways, as reductio arguments.

1) Goodman's grue (before and after t) - pertains to predictions.
2) Kripke's skeptical challenge to addition (bigger or smaller than 57) - play on Wittgenstein's rule following.
3) Fodor's comical take on the first, fridgeon (Fodor's fridge is on.) - demonstrates the frame problems.

We will focus on Kripke's version, as it goes deepest into the problem. It provides an interesting challenge to the primacy of Peircean diagrammatical reasoning.

Kripke formulated the problem to argue that the factual, hypothetical, or logical compatibilities are not enough to ground meaning: meaning is normative or at least non-factual (akin to Brandom's position).

I will briefly review the existing objections and alternatives this solution. However, my aim in this section is two-fold: 1) would the mechanism of an implicit definition help shed light on this discussion? The short answer is in the affirmative, the construction of the problem itself is an illegitimate move as per our previous construction of implicit definition. They are simply breaking the implicit equivalence on purpose by adding new conditions. And this provides the solution to the puzzles. 2) How would a modal

possibility of interaction fare? The preview of the answer is that it is complicated, simply posing the modal possibility of interaction boils down to the dispositionist position below, and as such would be insufficient. And yet, the historicity, agent detectable error as well as heuristics and algorithms may contribute to the realistically complex picture of the implementation of implicit definition. Norms of correctness as Exemplified heuristics or algorithms as actual flows, actual pathways of what works vs abstracted rules as types.

The distinction of dispositions versus normativities emerge again. Is the force of reasoning correctly, normative, or dispositional? (Recall section 3.2.1 on the conception of meaning as use). In posing this problem, Kripke's Wittgenstein is pointing to the inherent non-factuality involved in rule-following and meaning. *Non-factualism* views ascriptions of meaning as serving a purpose or function distinct from that of making factual statements. Brandom makes a similar point:

> core idea is that anything recognizable as an intentional state (for present purposes, we can think of these as propositional contentful states or as conceptually contentful representations) must underwrite normative assessments as to whether things are as they ought to be, according to that state-whether the state is correct or successful according to the standards determined by its content. (2001, p. 589); For what an organism is doing to be intelligible as representing, there must be room also for misrepresenting, for representation that is incorrect. One of the hallmarks of the normativity of intentionality is that what one commits oneself to in applying a concept outruns in principle what one takes oneself to be committed to. The norm of correctness one thereby binds oneself by goes beyond both the dispositions of those undertaking those commitments and what they consciously envisage themselves as committing themselves to thereby. (2001, p. 590)

However, there is a problem: sources of normativity may differ. For instance, normativity involved with pragmatic entrenchment of color recognition is different from

linguistic normativity. Here we see a parallel of what we saw in chapter 3, on the source of the force of the rule: is it merely a psychological disposition, or is it norms?

However, when considering a sentence in terms of its function, there is a meaningful distinction between using it correctly or incorrectly, indicating that the sentence follows certain rules. This raises the question of what gives these rules their authority or power.

Uttering the sentence 'SS' is deemed correct if the circumstances under which it is spoken align with the conditions set by rule R1.

Conditions play a role, but conditions are unclear (possibly also unlimited and uncertain), so the resulting categories should be implicit and opaque. There is a tension between non-factualism (a milder version of inferentialist normativity) and dispositionalism.

In reductive dispositionalism, the fact that Jones possesses the concept of addition instead of quaddition is identified with his disposition to produce the outcome of adding (rather than quadding) the numbers x and y in response to arithmetic queries such as “x + y =?”. Similarly, the fact that he attributes the meaning of "dog" to the word 'dog' is identified with his disposition to apply the term 'dog' to actual dogs.

Response to the non-factualist would be that the agents do not consciously envisage themselves, i.e., the bee dance. They run through the actual heuristic, they do not have to be aware of what the heuristic or the disposition is. However, the heuristic may implicitly differentiate, implicitly embody the norms of correctness (conditions) without the norms actually being present.

5.2 Systematicity, productivity, compositionality

Systematicity, productivity and compositionality are defining features of semantic phenomena, that apply to thought and language. When one rejects the computationalist (encodingist) view of "representational states with a compositional constituent structure" (Matthews, 2006), one needs to account for these features in a way other than compositionality.

The systematic nature of content was described by Gareth Evans, who formulated it as the generality principle in his Varieties of Reference (1982). In the book, he attempts to give a theoretical explanation of Russell's principle. In doing this, he criticizes verificationist accounts of semantics. And he postulates the generality constraint as a constraint (prerequisite) on thinkability:

> In discussing the nature of our conceivings we have little enough to go on, but there is one fundamental constraint that must be observed in all our reflections: I shall call it 'The Generality Constraint." (p. 100); if a subject can be credited with the thought that a is F, then he must have the conceptual resources for entertaining the thought that a is G, for every property of being G of which he has a conception. This is the condition that I call "The Generality Constraint". (p. 104)

Furthermore:

> We thus see the thought that a is F as lying at the intersection of two series of thoughts: on the one hand, the series of thoughts that a is F, that b is F, that c is F, [and so on], and, on the other hand, the series of thoughts that a is F, that a is G, and that a is H, [and so on]. (p.104, footnote 21)

Proponents of computationalist accounts such as Fodor define the productivity, systematicity and compositionality in terms of each other, as mutually supporting conditions of human cognitive systems, since their semantics is based on composition as structure, or compositional correspondence. In their definitions, systematicity and

productivity gets interrelated, as productivity seems to be the capacity to construct new content, by means of new combinations or variations, however this capacity is taken to be systematically bounded. There are other interpretations of systematicity, in a sense more closely akin to exhaustiveness,

They claim that the systematic relations between the contents of one's thoughts is evidence for the fact that the underlying representational structures and reasoning abilities of the agent must also be correspondingly systematic and productive, and this can only be achieved with compositionality of mental content[79], from which systematicity and productivity are by definition derivative. If one is operating with compositional correspondence, productivity requires decomposition, because new strings are formulated by recombination of decomposed atoms of meaning. Even though such productivity would provide unbounded number of elements, there could be no construction of new atomic elements, no true novelty or creativity.[80]

One must give a theoretical grounding of compositionality in order to use it to ground systematicity. Compositionality cannot simply be a self-appointed capacity. Syntactical computational compositionality is not a necessary criterion for human cognition, while generality/ systematicity is. One of the points we have argued in this thesis is that compositional (structural) correspondence is not the only kind of correspondence, because implicit defining is not compositional, and yet provides a

---

[79] Note that Gareth Evans assumed the generality constraint to apply to thought while not necessarily committing to compositionality and a language of thought.

[80] Furthermore, as Boden points out (Boden, 2004, p. 41), the problem with taking mere "surprising" combinations as creative is that, they lack criteria of meaning and use. The difficulty with creativity is always coming up with new criteria, without criteria about what makes something useful or meaningful, iterative or surprising combinations are not interesting. As a result, we cannot expect genuine creation from combination without additional criteria.

different, and more powerful kind of correspondence. What, then, could be the corresponding functional accounts of productivity and systematicity?

Productivity is resourcefulness, ability to produce further variations. Productivity is ubiquitous to all action systems. It was recognized as part of language and linguistic systems, but physically moving is productive, problem solving is productive. Productivity involves not only navigating but also creating new possibilities in the topology of possibilities. So, productivity can proceed by interacting, even with partial knowledge.

Use of (partial) concepts can be instrumental for generality. As for tracking, if meaning lies not in the atomistic tokens of mental particulars, but in the roles, these play within a system, so if the tracked is the functional roles instead of the intentional objects, then tracking becomes possible even with partial knowledge. In other words, if we take conceptual senses as abilities, there is a gradient of possibilities in which we can achieve this performance completely. For instance, just intending the thought is enough to render the act of thinking itself consummated without necessitating full epistemic accuracy.

Thus, is computational productivity being the exhaustive power to provide new combinations, functional productivity is creativity, an all-purpose, agile capacity to fulfill any goal criteria, transform any situation.

As for systematicity, structural systematicity may be exhaustiveness, functional systematicity is integration, unity. Systematicity and productivity are interlinked, defined in terms of each other. Systematicity concerns covering numerical possibilities exhaustively, the ability to. Thus, directly relevant to the usefulness or functionality of language, language limits the unlimited possible interpretations of a situation and allows for systematic consideration of options. Language allows limiting and framing, which is

not a given. Decreasing the possibilities is key for successful interaction and intervention, for example baby crying, or computer troubleshooting.

CHAPTER 6

CONCLUSION

We have covered some ground with regards to the changing nature of representing relationship. Some of the important conceptual distinctions that aid the exposition were the difference of exemplification versus abstraction of cognitive – inferential procedures ("maps" in Sellars' and Millikan's sense), the functional distinction between implicit and explicit, the derivation of extension from intension, cause as intervention, measure theory as a way to demonstrate the transformations involved in extracting representations, extracting particularly structures as invariances from interactions and interventions, and a revised notion of functional productivity (as opposed to combinatorial productivity) whereby new ways of interacting are produced.

There is much more to be said with regards to these concepts, each of them can be further developed to concretely address philosophical problems like the frame problems, the problem of creativity, logic of abduction, incommensurability of concepts, and so on. The aim of this thesis is expository, to connect these concepts and argue for the causal and ontological reality of pragmatist approaches, which is a foundational issue for the integrating of philosophy. Otherwise, we are stuck with parallel rhetorical moves in different discussions while missing the opportunity to apply them more generally.

REFERENCES

Allen, J. W. P., & Bickhard, M. H. (2011). You can't get there from here: Foundationalism and development. *Behavioral and Brain Sciences*, *34*(3), 124–125. https://doi.org/10.1017/S0140525X10002311

Andreas, J., Beguš, G., Bronstein, M. M., Diamant, R., Delaney, D., Gero, S., Goldwasser, S., Gruber, D. F., de Haas, S., Malkin, P., Payne, R., Petri, G., Rus, D., Sharma, P., Tchernov, D., Tønnesen, P., Torralba, A., Vogt, D., & Wood, R. J. (2021). *Cetacean Translation Initiative: A roadmap to deciphering the communication of sperm whales*. *arXiv:2104.08614*. https://doi.org/10.48550/arXiv.2104.08614

Bateson, G. (2000). *Steps to an ecology of mind: Collected essays in anthropology, psychiatry, evolution, and epistemology*. University of Chicago Press. Retrieved from https://press.uchicago.edu/ucp/books/book/chicago/S/bo3620295.html (Original work published 1973)

Bateson, G. (2015). Form, substance and difference. *ETC: A Review of General Semantics*, *72*(1), 90–104.

Behrmann, M., & Nishimura, M. (2010). Agnosias. *Wiley Interdisciplinary Reviews: Cognitive Science*, *1*, 203–213. https://doi.org/10.1002/wcs.42

Bickhard, M. H. (1980). *Cognition, convention, and communication*. Praeger.

Bickhard, M. H. (1989). The nature of psychopathology. In L. Simek-Downing (Ed.), In *International psychotherapy: Theories, research, and cross-cultural implications* (pp. 115–140). Praeger Publishers.

Bickhard, M. H. (1998). Levels of representationality. *Journal of Experimental & Theoretical Artificial Intelligence*, *10*(2), 179–215. https://doi.org/10.1080/095281398146824

Bickhard, M. H. (2009a). Interactivism: A manifesto. *New Ideas in Psychology*, *27*(1), 85–95. https://doi.org/10.1016/j.newideapsych.2008.05.001

Bickhard, M. H. (2009b). The interactivist model. *Synthese*, *166*(3), 547–591. https://doi.org/10.1007/s11229-008-9375-x

Bickhard, M. H. (2023). *The whole person* [Book Manuscript].

Bickhard, M. H., & Terveen, L. (1995). Foundational issues in artificial intelligence and cognitive science: Impasse and solution. In *Advances in Psychology, 109* (Vol.

109). Elsevier Science Publishers. https://psycnet.apa.org/record/1995-97788-000

Bitbol, M. (2007). Physical relations or functional relations. A non-metaphysical construal of Rovelli's relational quantum mechanics. [Preprint] Retrieved from http://philsci-archive.pitt.edu/3506/

Bitbol, M. (2012). Downward causation without foundations. *Synthese*, *185*(2), 233–255.

Block, N. (Ed.). (1980). *Readings in philosophy of psychology, Volume I:* Harvard University Press.

Boden, M. A. (2004). *The creative mind: Myths and mechanisms*. Routledge.

Brandom, R. (1994). *Making it explicit: Reasoning, representing, and discursive commitment*. Harvard University Press.

Brandom, R. (2001). Modality, normativity, and intentionality. *Philosophy and Phenomenological Research*, *63*(3), 587–609. https://doi.org/10.2307/3071157

Brentano, F. (2012). *Psychology from an empirical standpoint*. Routledge.

Brooks, R. A. (1991). Intelligence without representation. *Artificial Intelligence*, *47*(1), 139–159. https://doi.org/10.1016/0004-3702(91)90053-M

Campbell, D. T. (1960). Blind variation and selective retentions in creative thought as in other knowledge processes. *Psychological Review*, *67*, 380–400. https://doi.org/10.1037/h0040373

Campbell, D.T. (1974). Evolutionary epistemology. In P. A. Schilpp (Ed.), *The philosophy of Karl Popper*. Open Court Publishing.

Campbell, D. T. (1987). Evolutionary Epistemology. In G. Radnitzky & W. W. Bartley, III (Eds.), *Evolutionary Epistemology, Rationality, and the Sociology of Knowledge* (pp. 47–89). Open Court Publishing.

Campbell, R. J., & Bickhard, M. H. (2011). Physicalism, emergence and downward causation. *Axiomathes*, *21*(1), 33–56. https://doi.org/10.1007/s10516-010-9128-6

Carey, S. (2004). Bootstrapping & the origin of concepts. *Daedalus*. https://doi.org/10.1162/001152604772746701

Carey, S. (2009). *The origin of concepts*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195367638.001.0001

Carnap, R. (1936). Testability and meaning. *Philosophy of Science*, *3*(4), 419–471.

Carnap, R. (1967). *The logical structure of the world*. Berkeley-Los Angeles, Univ.

Carnap, R. (1988). *Meaning and necessity: A study in semantics and modal logic* (2nd Edition, Vol. 30). University of Chicago Press. (Original work published 1947)

Christopoulou, D. (2013). On the synthetic content of implicit definitions. *Logic and Logical Philosophy*, *22*(1), Article 1. https://doi.org/10.12775/LLP.2013.005

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*(3), 181–204. https://doi.org/10.1017/S0140525X12000477

Crane, T. (2001). *Elements of mind*. Oxford University Press.

Cummins, R. (1980). Functional analysis. In N. Block (Ed.), *Readings in philosophy of psychology, Volume I:* (pp. 185–190). Harvard University Press.

Currie, G. (1982). Frege's letters. *The British Journal for the Philosophy of Science*, *33*(1), 65–77.

Davis, R., Shrobe, H., & Szolovits, P. (1993). What Is a Knowledge Representation? *AI Magazine*, *14*(1), Article 1. https://doi.org/10.1609/aimag.v14i1.1029

Dennett, D. C. (1991). Granny's campaign for safe science. *Meaning in mind: Fodor and his critics*, 87–94.

Díez, J. A. (1997). A hundred years of numbers. A historical introduction to measurement theory 1887–1990: Part I: The formation period. Two lines of research: Axiomatics and real morphisms, scales and invariance. *Studies in History and Philosophy of Science Part A*, *28*(1), 167–185. https://doi.org/10.1016/S0039-3681(96)00014-3

Egan, F. (2014). How to think about mental content. *Philosophical Studies*, *170*(1), 115–135. https://doi.org/10.1007/s11098-013-0172-0

Emmeche, C., Køppe, S., & Stjernfelt, F. (2000). Levels, emergence, and three versions of downward causation. *Downward Causation. Minds, Bodies and Matter*, 13–34.

Evans, G. (1982). *The varieties of reference* (J. McDowell, Ed.). Oxford University Press.

Fodor, J. A. (1978). Tom Swift and his procedural grandmother. *Cognition*, *6*(3), 229–247.

Fodor, J. A. (2005). *Hume Variations*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199287338.001.0001

Freudenthal, H. (1966). The main trends in the foundations of geometry in the 19th century. In *Studies in Logic and the Foundations of Mathematics* (Vol. 44, pp. 613–621). Elsevier.

Gabriel, G. (1978). Implizite definitionen—eine verwechselungsgeschichte. *Annals of Science*, *35*(4), 419–423.

Galilei, G. (1612). *Discorso intorno alle cose che stanno in su l'aqua o che in quella si muovono In Opere I*. (Torino: U.T.E.T).

Gibson, J. J. (2015). *The ecological approach to visual perception (Classic edition)*. Psychology Press. (Original work published 1979)

Giovannini, E. N., & Schiemer, G. (2021). What are implicit definitions? *Erkenntnis*, *86*(6), 1661–1691. https://doi.org/10.1007/s10670-019-00176-5

Glymour, C., & Eberhardt, F. (2022). Hans Reichenbach. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2022). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/spr2022/entries/reichenbach/

Goodman, N. (1983). *Fact, fiction, and forecast*. Harvard University Press.

Goodman, N., & Elgin, C. Z. (1988). *Reconceptions in philosophy and other arts and sciences*. Hackett Publishing.

Gupta, A. (2011). *Truth, meaning, experience*. Oxford University Press.

Harnad, S. (1990). The symbol grounding problem. *Physica D: Nonlinear Phenomena*, *42*(1), 335–346. https://doi.org/10.1016/0167-2789(90)90087-6

Hempel, C. G. (1951). The concept of cognitive significance: A reconsideration. *Proceedings of the American Academy of Arts and Sciences*, *80*(1), 61–77. https://doi.org/10.2307/20023635

Hilbert, D. (1902). *The Foundations of Geometry. Authorised translation by EJ Townsend* (E. J. Townsend, Trans.). Kegan Paul & Company.

Jenn. (2021, February 8). *Bijection* [MOOC]. Calcworkshop. https://calcworkshop.com/functions/bijection/

Jones, M., & Love, B. C. (2011). Bayesian fundamentalism or Enlightenment? On the explanatory status and theoretical contributions of Bayesian models of cognition. *Behavioral and Brain Sciences*, *34*(4), 169–188. https://doi.org/10.1017/S0140525X10003134

Kalman, R. E. (2022). Control theory. In *Encyclopedia Britannica*. https://www.britannica.com/science/control-theory-mathematics

Kirsh, D. (1991). Today the earwig, tomorrow man? *Artificial Intelligence*, *47*(1), 161–184. https://doi.org/10.1016/0004-3702(91)90054-N

Kirsh, D. (2006). Implicit and explicit representation. In *Encyclopedia of Cognitive Science*. John Wiley & Sons, Ltd. https://doi.org/10.1002/0470018860.s00166

Kirsh, D. (2009). Interaction, External Representation and Sense Making. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *31*(31), 1069-7977. https://escholarship.org/uc/item/9rg717jz

Kirsh, D. (2011). How marking in dance constitutes thinking with the body. *Versus: quaderni di studi semiotici, (112), 183-214.*

Levesque, H. J., & Brachman, R. J. (1987). Expressiveness and tractability in knowledge representation and reasoning. *Computational Intelligence*, *3*(1), 78–93. https://doi.org/10.1111/j.1467-8640.1987.tb00176.x

Malt, B. C. (1994). Water is not H2O. *Cognitive Psychology*, *27*(1), 41–70.

Matthews, R. J. (2006). Connectionism and systematicity. In *Encyclopedia of Cognitive Science*. John Wiley & Sons. https://doi.org/10.1002/0470018860.s00164

Matthews, R. J. (2010). *The measure of mind: Propositional attitudes and their attribution*. Clarendon Press.

Millikan, R. G. (1989). Biosemantics. *The Journal of Philosophy*, *86*(6), 281–297. https://doi.org/10.2307/2027123

Millikan, R. G. (2005). The father, the son, and the daughter: Sellars, Brandom, and Millikan. *Pragmatics & Cognition*, *13*(1), 59–71. https://doi.org/10.1075/pc.13.1.06mil

Moran, D. (2014). Descartes on the formal reality, objective reality, and material falsity of ideas: Realism through constructivism? In *Realism, science, and pragmatism* (pp. 75–100). Routledge.

Orłowska, E., & Pawlak, Z. (1984). Expressive power of knowledge representation systems. *International Journal of Man-Machine Studies*, *20*(5), 485–500. https://doi.org/10.1016/S0020-7373(84)80023-1

Peirce, C. S. (1994). *The collected papers of Charles Sanders Peirce. Electronic edition* (J. Deely, Ed.). Harvard University Press.

Piaget, J. (1970). *Genetic epistemology.* New York Chichester, West Sussex: Columbia University Press. https://doi.org/10.7312/piag91272

Poggiolesi, F. (2012). Conservativeness and eliminability for anti-realistic definitions. In S. Rahman, G. Primiero, & M. Marion (Eds.), *The realism-antirealism debate in*

the age of alternative logics (pp. 169–180). Springer Netherlands. https://doi.org/10.1007/978-94-007-1923-1_9

Poincare, H. (2012). *The value of science: Essential writings of Henri Poincare*. Random House Publishing Group. (Original work published 1902)

Popper, K. R. (1964). *Creative and non-creative definitions in the calculus of probability*. Springer.

Popper, K.R. (1965). *Conjectures and refutations*. New York: Harper & Row.

Putnam, H. (1980). The nature of mental states. In Ned Block (Ed.), *Readings in philosophy of psychology: Vol 1* (pp. 223–231). Harvard University Press.

Quine, W. V. O. (1964). Implicit definition sustained. *The Journal of Philosophy*, *61*(2), 71–74. https://doi.org/10.2307/2023271

Quine, W. V. O. (2013). *Word and object*. MIT press. (Original work published 1960)

Quine, W. V. O. (1953). Two dogmas of empiricism. In Quine, W. V. O., *From a Logical Point of View*. Harvard University Press.

Reck, E. H., & Price, M. P. (2000). Structures and structuralism in contemporary philosophy of mathematics. *Synthese*, *125*(3), 341–383. https://doi.org/10.1023/A:1005203923553

Rescher, N. (1973). Conceptual idealism. Oxford: Blackwell.

Rescher, N. (1991). Conceptual idealism revisited. *The Review of Metaphysics*, *44*(3), 495–523.

Resnik, M. D. (1974). The Frege-Hilbert controversy. *Philosophy and Phenomenological Research*, *34*(3), 386–403.

Rosenthal, S. B. (1994). *Charles Peirce's pragmatic pluralism*. State University of New York Press.

Schlick, M. (1974). *General theory of knowledge*. Springer Vienna. https://doi.org/10.1007/978-3-7091-3099-5

Searle, J. R. (1984). *Minds, brains and science*. Harvard University Press.

Sellars, W. (1953a). Inference and meaning. *Mind*, *62*(247), 313–338.

Sellars, W. (1953b). Is there a synthetic a priori? *Philosophy of Science*, *20*(2), 121–138.

Sellars, W. (1963). Some reflections on language games. In *Science, perception and reality* (pp. 323–358). Ridgeview Publishing Company.

Sellars, W. (1974). Meaning as functional classification. *Synthese*, *27*(3), 417–437. https://doi.org/10.1007/BF00484606

Sellars, W. (1979). *Naturalism and ontology*. Ridgeview Reseda.

Sellars, W. (1997). *Empiricism and the philosophy of mind* (R. Rorty & R. B. Brandom, Eds.). Harvard University Press. (Original work published 1956)

Shapiro, S. (1997). *Philosophy of mathematics: Structure and ontology*. Oxford University Press.

Shea, N. (2018). *Representation in cognitive science*. Oxford University Press.

Sloman, A. (2018). What did Bateson mean when he wrote" information" is" a difference that makes a difference"? *School of Computer Science, The University of Birmingham, UK*. http://www.cs.bham.ac.uk/research/projects/cogaff/misc/information-difference.html

Strogatz, S. (2019). *Infinite powers: How calculus reveals the secrets of the universe*. Houghton Mifflin Harcourt.

Suppes, P. (2002). *Representation and invariance of scientific structures*. CSLI Publications.

Swoyer, C. (1991). Structural representation and surrogative reasoning. *Synthese*, *87*(3), 449–508. https://doi.org/10.1007/BF00499820

Tarski, A., Mostowski, A., & Robinson, R. M. (1953). *Undecidable theories*. North-Holland Publishing Company.

Thorpe, L. (2021, June 25). *Pluralistic teleosemantics: Why we need both Bickhard Representations and Millikan Representations* [Conference presentation]. *Interactivist Summer Institute 2021*. Interactivism in Perspective, online. https://youtu.be/BhpiT-l_Tno

Uebel, T. (2016). Rudolf Carnap: The logical structure of the world. In J. Shand (Ed.), *Central works of philosophy Vol 4* (pp. 129–151). Routledge.

Van Gelder, T. (1995). What might cognition be, if not computation? *The Journal of Philosophy*, *92*(7), 345–381. https://doi.org/10.2307/2941061

Wilson, F. (1968). Definition and discovery. *The British Journal for the Philosophy of Science*, *18*(4), 287–303.

Wilson, F. (2013). *Acquaintance, ontology, and knowledge* (Vol. 19). De Gruyter. https://doi.org/10.1515/9783110327014

Wittgenstein, L. (1922). *Tractatus logico-philosophicus: German and English* (C. K. Ogden, Trans.). Routledge. https://www.routledge.com/Tractatus-Logico-Philosophicus-German-and-English/Wittgenstein/p/book/9780415051866

Wittgenstein, L. (1953). *Philosophical investigations* (G. E. M. Anscombe, Trans.). Oxford-Blackwell.

Younghouse, P. (1982, Spring). [Review of the book *Cognition, convention, and communication,* by M. Bickhard]. *Journal of Visual Verbal Languaging, 55-57*. https://www.tandfonline.com/doi/abs/10.1080/23796529.1982.11674350

Zhuli. (2023, March 29). *The most important counting concept you've (probably) never heard of [Video]*. YouTube. https://youtu.be/uxjAHkn4VFI