Doctoral Dissertation

Real Pattern Emergence: A Study of Patternhood, Generality, and Emergent Behavior from a Peircean Perspective

March, 2021

Osaka University Graduate School of Human Sciences, Department of Human Sciences, Future Innovation Kyosei Studies, Philosophy of Kyosei

Jimmy J. Aames

Acknowledgements

I have received considerable support from a number of individuals throughout the writing of this dissertation. First of all, I want to thank my supervisor Tatsuya Higaki, not only for his advice at various stages of writing this dissertation, but also for introducing me to the fascinating world of philosophy when I was an undergraduate student, and guiding me through the graduate program at Osaka University. This dissertation would not have been possible without his thoughtful assistance throughout the years.

It is with equal pleasure that I express my gratitude to Kunihisa Morita, Masato Ishida, and Nobuo Fujikawa, who kindly offered their precious time to review my work and serve on my dissertation committee. I have benefitted greatly from the insightful feedback I received during Prof. Morita's philosophy of science seminars, and from the stimulating conversations on Peirce I have had with Prof. Ishida. Very special thanks go to André De Tienne, who guided my studies while I was at Indiana University-Purdue University Indianapolis and initiated me into the labyrinthine world of Peirce's thought, and to Yoriyuki Yamagata, whose perceptive comments in discussions on sundry topics of philosophy have been most invaluable. I also want to thank the Japan Society for the Promotion of Science (JSPS) for their financial support: this work has been supported by JSPS KAKENHI Grant Number JP18J20438.

Finally, I owe more than I can express in words to my family. My deepest thanks go to my mother, who has supported and encouraged me in all of my pursuits, and to my fiancé Minami, for her understanding and love over the years. I would not be where I am today without her dedicated support.

Contents

Introduction 1

Chapter 1: Real Patterns 7
1.1 Dennett's Theory (1): Algorithmic Compressibility 7
1.2 Dennett's Theory (2): Predictive Power 9
1.3 Ladyman and Ross' Elaboration: Non-Redundancy 11
1.4 Multiple Instantiability 15

Chapter 2: Peirce's Theory of Generality 18
2.1 The Problem of Universals 19
2.2 Peirce's Realism in the "New List of Categories" 23
2.3 The Berkeley Review 27
2.4 The Pragmatic Maxim 32
2.5 Peirce's Late Argument for Scholastic Realism 37

Chapter 3: Real Pattern Emergence 40
3.1 Epistemological and Ontological Emergence 40
3.2 The Notion of Levels 45
3.3 Real Pattern Emergence 49
3.4 Bedau on Weak Emergence 52

Chapter 4: Downward Causation and Teleology 60
4.1 Mutual Entrainment and the "Virtual Governor" 61
4.2 Problems with Downward Causation 64
4.3 Downward Causation and RP Emergence 68
4.4 Peircean Teleology 72

Concluding Remarks 80

Supplementary Chapter: Peirce's "New List of Categories" 83
S.1 Opening sections (§§1–2) 84
S.2 Substance (§3) 87
S.3 Being (§4) 89
S.4 The three modes of mental separation (§5) 91
S.5 Method of deriving the categories (§6) 94
S.6 Quality/reference to a ground (§7) 97
S.7 Relation/reference to a correlate (§8): preliminary discussion 98
S.8 Relations of equiparance and disquiparance 99
S.9 De Tienne's interpretation of the correlate 101
S.10 My interpretation of the correlate 102
S.11 The notion of comparison 105
S.12 Representation/reference to an interpretant (§9) 106
S.13 The double function of the interpretant 109
S.14 The reversal of Kant (§10) 110
S.15 The list of categories (§§11–13) 112
S.16 Regularity as the basis of cognizability 114

Abbreviations 117
Bibliography 117

Introduction

We live in a world of patterns. Slightly modifying Charles S. Peirce's (1839–1914) famous statement about signs (EP 2:394, 1906),[1] we can say that this universe is perfused with patterns, if not composed exclusively of patterns. Indeed, exhibiting patternhood seems to be a necessary condition for the very possibility of cognition. If to cognize something or perceive something involves recognizing a certain pattern, then it would seem that anything that does not exhibit any kind of patternhood is *ipso facto* incognizable. This is not to deny that there are purely random events, or at least events that appear to be so, such as the throw of a die or the radioactive decay of a particular atom. But such random events too display statistical regularities when they occur in large numbers, and moreover, we are able to perceive these random events only upon the backdrop of a series of regularities and invariants—we can perceive the die, for example, because it maintains a certain shape and rigidity, and does not, say, randomly evaporate into thin air.

But what is a *pattern?* It is not a *thing* in the usual sense of the word, but rather a regularity displayed by a series of things or events. And yet there are patterns that seem to be *real*, in the sense that by discerning them we are able to make predictions about future events that have a good chance of being fulfilled—laws of nature and personalities are prime examples of such patterns. But how can something that is not a *thing* nonetheless be real? What kind of ontological status do patterns have? Furthermore, what kind of relation do they have with the particular elements that constitute them? These are the questions with which this dissertation will be concerned. While our point of departure will be Daniel Dennett's theory of real patterns (Dennett 1991), the philosophy of Peirce—in particular his pragmatism and Scholastic realism—will serve as a guiding framework throughout this work.

[1] See the list at the end of this dissertation for an explanation of the abbreviations used in referring to Peirce's writings.

I have previously dealt with the topic of patterns in my Master's Thesis (Aames 2016). There, I attempted to formulate a general theory of patterns by comparing Peirce's philosophy with Ontic Structural Realism (OSR), a theory that is recently gaining attention in philosophy of science, especially philosophy of physics, and which claims that only patterns are real (individual objects are not real, or have only a "thin" being in some sense). Specifically, I dealt with the version of OSR developed by James Ladyman and Don Ross in their book *Every Thing Must Go* (Ladyman & Ross 2007). In this dissertation I want to approach the same topic that I addressed in my Master's Thesis from a slightly different angle. Instead of motivating the development of my theory through a discussion of OSR, here I want to highlight the various problems associated with the notion of *emergence*, and propose a new kind of emergence that I call *real pattern emergence* (or RP emergence for short). My aim will be to shed light on the nature of the autonomy that some patterns—namely, those which will be called *emergent real patterns* in this dissertation—possess with respect to their constituent elements.

This dissertation will be organized as follows. The first chapter will be devoted to an elucidation of the key concept of *pattern*. Taking Dennett's theory of patterns as our starting point, we will discuss the notion of algorithmic compressibility, which will serve as a criterion for when a pattern is present in some data, and then we will attempt to delineate the features that distinguish *real patterns* from *non-real patterns*. Ladyman and Ross' (2007) elaboration of Dennett's theory will also be addressed in this connection. Finally, I will discuss a key feature of patterns which I believe has not been sufficiently appreciated in the literature—namely, their *multiple instantiability*, or capacity of being instantiated by different particular elements. A melody played on the piano, for example, will retain its identity even if each of its individual notes is raised or lowered an octave. I will show that the multiple instantiability of patterns implies that they are *generals* (more commonly known as *universals*), and hence the question of their reality is seen to be a variation on the problem of universals, a central issue in the history of Western philosophy.

The generality of patterns makes them perfect candidates for applying Peirce's rich theory of generals, as encapsulated in his pragmatism and Scholastic realism. Throughout his life Peirce was a staunch defender of Scholastic realism, the doctrine that there are real generals, and not only did he marshal forth original arguments in support of this view, but he also *redefined* the traditional problem of universals in such a way that makes it more pertinent to the modern scientific worldview, while at the same time preserving the essence of the issue at stake. In Chapter 2 we will explore Peirce's unique brand of Scholastic realism, and how it is intimately tied together with his pragmatism.

Particular attention will be paid to a distinction that Peirce draws in his later works, between what he calls *will-be's* and *would-be's*. These terms refer to distinct types of modality. A *will-be* is a statement about what will happen in a given kind of situation, whereas a *would-be* is a habit or tendency that dictates not only what *will* happen, but also what *would* happen in hypothetical situations that are never actualized. A central tenet of Peirce's philosophy is that generals have the modality of *would-be's*: ascribing a general to some particular object x is equivalent to recognizing that x is governed by a series of laws or regularities that tell us not only how x *will* behave, but also how it *would* behave in certain kinds of counterfactual situations. This distinction between *will-be's* and *would-be's* will play a key role in elucidating the nature of the autonomy characteristic of emergent real patterns in the following chapter.

In Chapter 3 we will take up the issue of emergence and outline the concept of RP emergence. A distinction often drawn in the literature on emergence is that between *epistemological emergence* (also called *weak emergence* or *conservative emergence*) and *ontological emergence* (also called *strong emergence* or *radical emergence*). Epistemological emergence is a kind of emergence that is only in the eyes of the beholder. An epistemologically emergent phenomenon is one that can in principle be reduced to—predicted or derived from—its underlying elements, but is in practice irreducible to these elements due to epistemic limitations on the part of the observer (limitations in computational resources, knowledge about initial conditions, etc.). An on-

tologically emergent phenomenon, on the other hand, is one that is irreducible to its underlying elements even in principle.

Let us suppose that there are no instances of ontological emergence in the world, and that every emergent feature ultimately emerges epistemologically from the laws and entities of fundamental physics. This view, which I will call *ontological reductionism*, is widely endorsed by both philosophers and scientists. However, as I will argue in Chapter 3, this view is problematic, because it fails to recognize that the patterns we observe at everyday scales, as well as those studied by the various special sciences, are just as real and fundamental as the patterns studied by fundamental physics. On the other hand, the view that there are instances of ontological emergence in the world—a view which I will call *radical emergentism*—is also problematic, in that it posits something in the world that can only be described as sheer magic, something that simply pops into existence without any why or wherefore. My aim in this chapter will be to introduce the concept of RP emergence, demonstrate how it is distinct from both epistemological and ontological emergence, and show that many putative cases of emergent phenomena are actually examples of RP emergence. We will thus be able to avoid both ontological reductionism and radical emergentism, since if there is a third type of emergence that is neither epistemological nor ontological, then denying the existence of ontological emergence will not entail that all instances of emergence in the world are merely epistemological.

The real pattern that emerges in any instance of RP emergence is of such nature that it supports predictions about not only what *will* happen in a given situation, but also what *would* happen in an indefinite variety of possible micro situations (i.e., possible states and arrangements of the pattern's constituent elements) that are never actualized. Emergent real patterns thus carry more information than descriptions of their underlying elements and processes, making them *autonomous* in a strong sense from their emergent base. And as I hope to show, this autonomy is what differentiates RP emergence from merely epistemological emergence, without, on the other hand, collapsing it into ontological emergence. Further, I will try to throw the concept of RP

emergence into sharper relief by comparing it with Mark A. Bedau's (1997, 2002) related notion of *weak emergence* (not to be confused with epistemological emergence).

Finally, in Chapter 4, I will take up the issue of downward causation, often discussed in connection with ontological forms of emergence, and show how RP emergence may accommodate a teleological form of downward causation. After motivating the idea of downward causation through a discussion of the phenomenon of *mutual entrainment*, I will address two problems that seem to be inherent in the concept of downward causation, which I call the *incoherence problem* and *dispensability problem*, respectively. I will then argue that these problems do not arise if we conceive of the downward cause—the entity that exerts a downward causal influence—as an emergent real pattern. Furthermore, I will try to show that this causal influence is best understood as a form of final causation as conceived by Peirce.

I have also added to this dissertation a supplementary chapter in which I provide a commentary on the first thirteen sections of Peirce's early paper "On a New List Categories" (1867). While the "New List" is discussed explicitly only in Chapter 2 of this dissertation, the theory of categories that Peirce first sets out in this seminal paper constitutes the undercurrent of many of the ideas developed in this dissertation. In particular, we shall see how the theory of cognition that Peirce presents in this paper implies that regularity is the basis of cognizability, as suggested in the opening of this introduction.

Given the rapid developments in the study of complex systems and self-organizing phenomena, as well as the advent of new techniques in machine learning, the concepts of patternhood and emergence are taking on a new significance in today's science. And yet these concepts have so far eluded attempts to place them on a firm philosophical basis. What I hope to do in this dissertation is to show that Peirce's philosophy of generality offers us a novel perspective on the issues surrounding patternhood and emergence, while at the same time updating Peirce's Scholastic realism in the form of a realism about patterns. This, I believe, will provide a demonstra-

tion of the abiding relevance of Peirce's philosophy in thinking about foundational issues in today's science.

Chapter 1: Real Patterns

> *[M]ere individual existence or actuality without any regularity whatever is a nullity. Chaos is pure nothing.*
> Charles S. Peirce, "What Pragmatism Is"

In this chapter we will study the general properties of patterns, taking Dennett's (1991) theory of patterns as our point of departure. I will give a definition of *pattern* in terms of the notion of algorithmic compressibility (§1.1) and examine Dennett's distinction between real and non-real patterns in terms of their predictive power (§1.2). I will then take up Ladyman & Ross' (2007) elaboration of Dennett's theory (§1.3), and finally I will argue that patterns are *general* in the sense of being instantiable by different particular elements (§1.4).

§1.1 Dennett's Theory (1): Algorithmic Compressibility

To begin with, it will be useful to have a definition of the key term *pattern*. Here I will be relying on the theory of real patterns proposed by Dennett in his paper "Real Patterns" (Dennett 1991). In the most general terms, a pattern is a regularity in some data, where *data* is construed in the broadest possible sense as something that is observed or may be observed. Consider, for example, an endless random string of 0's and 1's. There is no regularity in this data. On the other hand, consider an endless string of alternating 0's and 1's: 010101010 … etc. What we should notice is that this data can be compressed into a program that commands: "generate an endless string of alternating 0's and 1's." There is no way of compressing the random string of 0's and 1's—the only way this data can be transmitted to another person is to send the *bit map*, which identifies each digit *seriatim* (the first place value is 0, the second place value is 0, the third place value is 1, etc.). In more general terms, a bit map is a zero-compression encoding, where each bit of information in the initial data is mapped one-to-one to a distinct bit in the encoding.

Gregory Chaitin, one of the founders of algorithmic information theory, gives the following definition of *randomness*: "A series of numbers is random if the smallest algorithm capable of specifying it to a computer has about the same number of bits of information as the series itself" (Chaitin 1975:48). Reversing this idea, Dennett proposes the following criterion for the presence of a pattern: "A pattern exists in some data—is real—if there is a description of the data that is more efficient than the bit map, whether or not anyone can concoct it" (Dennett 1991:34). That is, there is a pattern in some data if there is an algorithm that reproduces the data using a smaller number of bits than the data itself (when there is such an algorithm, we say that the data is *algorithmically compressible*).[2]

An interesting aspect of pattern recognition is that not all observers are able to discern the same pattern in the same data, and even the same observer may discern different patterns in the same data on different occasions. The famous duck-rabbit illusion is a prime example of the latter. As an example of the former, suppose that an image file—say a jpg image of a human face—is translated into binary notation, pixel by pixel. The pattern is still there, but it would be impossible for the human eye to discern it visually. Other creatures with different sense organs may readily perceive patterns that are imperceptible to us (Dennett 1991:34). Hence Dennett's proviso that the presence of a pattern should not depend on whether or not anyone is actually able to concoct a compression algorithm: there is a pattern in some data if the data is *in principle* compressible by a potential observer.

[2] As we will see below, in the same paper Dennett gives another criterion for the reality of patterns, according to which a pattern is real if by discerning it one can make successful predictions about future events. The relation between this predictive power criterion and the algorithmic compressibility criterion cited here, however, is unclear. It seems to me that this unclarity derives from Dennett's ambiguous use of the term "real" in the quoted passage. The unclarity can be resolved if we take Dennett's algorithmic compressibility criterion as a criterion simply for the *presence* of a pattern, regardless of whether or not it is real (in the sense of supporting predictions about future events). I will touch upon this point again below.

§1.2 Dennett's Theory (2): Predictive Power

Dennett's algorithmic compressibility criterion specifies a necessary and sufficient condition for the *presence* of a pattern, but it does not by itself guarantee that the pattern is *real*. It is true that he formulates the criterion using the term "real": "A pattern exists in some data—is real—if there is a description of the data …" (Dennett 1991:34). However, I believe this is simply due to his ambiguous use of the word "real" in this paper, for he is clear throughout that there are non-real as well as real patterns, and he goes on to specify a further criterion for distinguishing between the two.

To get an idea of what a real pattern is, let us first consider what a non-real pattern is. A non-real pattern is one that is due to pure chance. Suppose, for example, that we are to throw an unbiased die one thousand times and record the outcome of each trial. Let us further suppose that after having done ten throws, we find that in all of the first ten throws the die turns up six. This, of course, is a pure accident, but if we were to show the results of the first ten trials to someone without telling her that the numbers were generated by throwing a die, and we were to ask her to predict the next number, she would most likely predict that it will also be a six. She may happen to be right, but in that case we ask her to predict the next number, and the next, and so on. Then her prediction is bound to fail sooner or later. A pattern is undoubtedly present in the results of the first ten trials, and indeed, the data for the first ten trials can be compressed into a program that commands: "generate ten sixes." Nonetheless, further accumulation of data will eventually reveal the accidental nature of this pattern.

Conversely, we can say that a pattern is real if by discerning it we are able to make better-than-chance predictions about future events. A real pattern is one that is projectible into unobserved instances. Or as Dennett puts it, a pattern is real if you can get rich by betting on it (Dennett 1991:36). Dennett himself is not altogether clear on the relation between this predictive power criterion and the algorithmic compressibility criterion. The two criteria are clearly not equivalent, as can be seen in the above example of throwing a die. There will be no obscurity,

however, if we simply take the predictive power criterion as a necessary and sufficient condition for the *reality* of a pattern, and the algorithmic compressibility criterion as a necessary and sufficient condition for the *presence* of a pattern, regardless of whether it is real or not. As we shall see in §2.5, Dennett's idea that the reality of patterns consists in their predictive power is strikingly similar to one of Peirce's late arguments for Scholastic realism.

The notion of predictive power brings us back to our earlier consideration, that not all observers are able to discern the same pattern in the same data, and that even the same observer may discern different patterns in the same data on different occasions. This means that patterns are in some sense *observer-dependent*. Dennett explicates this notion in terms of prediction: patterns are observer-dependent in that they can be discerned only from the point of view of an observer that adopts a certain predictive strategy, or *stance*, to use Dennett's terminology. For instance, Dennett calls the predictive strategy from which intentional states—beliefs, desires, and the like—can be discerned the *intentional stance* (Dennett 1987:17). Likewise, there can be predictive strategies for discerning any kind of pattern whatsoever: the Newtonian mechanics stance, the cellular biology stance, the microeconomic stance, etc. The idea is that patterns are not simply "out there," naked in the world; on the contrary, the recognition of a pattern must always involve an element of active participation on the part of the observer, namely the adoption of a certain predictive strategy. This should not be taken to mean that the act of adopting a predictive strategy is always a conscious, deliberate act: the decision of which predictive strategy to adopt is dictated to a large degree by the structure of our sense organs, our genetic makeup, and the evolutionary history of our culture (Dennett 1991:36). Patterns are also observer-dependent in the further sense that a pattern, by definition, must be a candidate for pattern *recognition* (Dennett 1991:32); an incognizable pattern is a contradiction in terms. This implies that there must be (at least potentially) someone or something that is capable of doing the recognizing.

Patterns thus have an observer-dependent being; but at the same time, they are in another sense *observer-independent*. They are observer-independent in that the facts about the success or failure of our predictive strategies do not depend on what we may think or will them to be; they are completely out of our control. It is this uncontrollability of the outcome of our predictions that imparts to some patterns—namely, those whose discernment leads to successful predictions—a real being.

§1.3 Ladyman and Ross' Elaboration: Non-redundancy

As mentioned in the introduction, Ladyman and Ross develop a version of Ontic Structural Realism (OSR) in their book *Everything Must Go* (Ladyman & Ross 2007). Recall that the central claim of OSR is that only patterns are real; individual objects are not real, or have only a "thin" being in some sense. Ladyman and Ross' version of OSR is based on an elaboration of Dennett's theory of real patterns. Here I will be concerned not so much with their version of OSR, as with their elaboration of Dennett's theory. In particular, the notion that they articulate of the *non-redundancy* of real patterns will have bearings on our discussion of RP emergence in Chapter 3.

While OSR will not be the main focus of what follows, a few remarks are in order regarding the claims of Ladyman and Ross' version of OSR. Ladyman and Ross do not deny the reality of everyday objects like tables and chairs, nor the objects studied by the special sciences; what they deny is that they are individuals.[3] What we traditionally conceive as individual "things" are reconceived as real patterns, discernable at certain grains of observational resolution: "Some real patterns … behave like things, traditionally conceived, while others behave like traditional instances

[3] For our purposes, an *individual* may be understood roughly as anything capable of subsisting by itself, independently of its relation with anything else. The notion of individuality, however, is hard to pin down exactly. In fact, OSR's rejection of individuals is ultimately based on the observation that this notion is fundamentally ambiguous: "It is an *ersatz* form of realism that recommends belief in the existence of entities that have such ambiguous metaphysical status" (Ladyman 1998: 420).

of events and processes" (Ladyman & Ross 2007:121). Of course, a non-OSRist may agree with this, and yet hold that reality "bottoms out" at some fundamental level of individual objects, such as the level of elementary particles. Indeed, as Ladyman and Ross (2007:1–7, 17–27) point out, many philosophers seem to take it for granted that this is the correct way of looking at the world. The radicalness of OSR consists in its claim that reality does not "bottom out" at such a fundamental level. As Ladyman and Ross put it, "it's real patterns all the way down" (Ladyman & Ross 2007:228).[4]

With these clarifications in place, let us turn to Ladyman and Ross' definition of real patterns, and see how they elaborate on Dennett's theory. Ladyman and Ross's main complaint against Dennett's criterion for the reality of patterns is that it is not stringent enough—on their view, the facilitation of successful predictions specifies a *necessary* condition for the reality of a pattern, but not a *sufficient* condition. They refer to an idea developed by Dennett in one of his early works on philosophy of mind (Dennett 1971), concerning the *indispensability* of the intentional stance in making predictions about certain systems. For example, it is *possible* for someone to assume the intentional stance to predict the behavior of a thermostat—"It prefers the room to be 68 degrees and believes it is now 64 degrees, so it decides to turn on the furnace"—but one *must* assume the intentional stance towards a chess-playing computer, in order to not lose predictive power (Ladyman and Ross 2007:206). In the case of the thermostat, the intentional stance is possible but dispensable, whereas in the case of the chess-playing computer, the intentional stance is *indispensable*: if one were to dispense with the intentional mode of data compression, then they would find it far more difficult—perhaps even impossible—to predict the computer's next move.

But the question is: indispensable for whom? Ladyman and Ross accuse Dennett of suggesting in his real patterns paper that indispensability should be relativized to a given level of error toler-

[4] In fact, Ladyman and Ross reject the very idea that there are "levels of reality," on the grounds that talk of "levels" is a metaphor to which current science gives no interesting content (Ladyman & Ross 2007:53–57); we will return to this issue in §3.2.

ance on the part of the observer—this, they argue, is too instrumentalist (Ladyman & Ross 2007:206). It seems to entail that there is no indispensability condition at all, since any mode of compression is presumably indispensable at *some* level of error tolerance. According to Ladyman and Ross, the indispensability of a mode of compression ought not to be relativized to the computational capacity of some arbitrarily distinguished computers in some arbitrarily limited observational circumstances, such as a group of humans (Ladyman and Ross 2007:208); otherwise we would fall into instrumentalism. Rather, the sufficient condition for the reality of a pattern should be the indispensability of the associated mode of compression by *any physically possible computer* (Ladyman & Ross 2007:221). Whether a given computation is physically possible can be determined by calculating the lower bounds of the energy required to carry out that computation, using a principle known as Landauer's principle (Ladyman & Ross 2007:208). A physically possible computer can then be defined as a device that only carries out physically possible computations. According to Ladyman and Ross, it is only by introducing this condition that we can make sense of Dennett's claim that there are real patterns that no person has yet discovered, or will ever discover, encapsulated in his proviso "whether or not anyone can concoct [a compression algorithm]" in his formulation of the criterion for the presence of a pattern (see §1.1 above).

On the basis of these considerations, Ladyman and Ross formulate their definition of real patterns as follows. A pattern is real iff:

> (i) it is projectible; and
>
> (ii) it has a model that carries information about at least one pattern P in an encoding that has logical depth less than the bit-map encoding of P, and where P is not projectible by a physically possible device computing information about another real pattern of lower logical depth than [the pattern at hand]. (Ladyman & Ross 2007:233)

To say that a pattern is *projectible* is to say that it is generalizable to unobserved cases; that is, one could make better-than-chance predictions about unobserved phenomena on the basis of a model or simulation of that pattern. Clause (i) is thus a restatement of Dennett's predictive power criterion for the reality of a pattern. The first half of clause (ii) states that the pattern should

encode information about another pattern P in an encoding that is more efficient than the bit-map encoding of P. *Logical depth*, a notion originally formulated by Charles H. Bennett, is a measure of complexity that Ladyman and Ross introduce into their definition of real patterns in order to give rigorous expression to the ideas of informational efficiency and non-redundancy. The logical depth of a given object is defined as "the time required by a standard universal Turing machine to generate [the object] from an input that is algorithmically random" (Bennett 1988:227). Finally, the second half of clause (ii) states that further compression of the pattern should be physically impossible without sacrificing its projectibility with respect to P. The idea expressed here is that in order for a pattern to be real, it should be informationally *non-redundant*—that is, it should be indispensable in making predictions about unobserved phenomena.

I fully agree with Ladyman and Ross that a non-redundancy criterion should be introduced in characterizing the reality of patterns. Such a criterion is necessary in order to rule out, for example, the "intentional" behavior of a thermostat from being counted as a real pattern. However, I will not endorse Ladyman and Ross' definition of real patterns, because I find their appeal to "physically possible" computers problematic. The problem is that appealing to physically possible computers is just as arbitrary as appealing to a specific group of computing agents such as humans. Suppose that we determine whether a given computation is physically possible by using Landauer's principle to calculate the lower bounds of the energy required to carry out that computation, and then checking whether this lower bound exceeds the total amount of energy available in the universe. But we do not know the total energy of the universe, and it is quite possible that it is zero (Pasachoff & Filippenko 2019:613). Perhaps we could instead use the total amount of energy available in a particular region of the universe, such as our solar system. But why should the ontological status of a pattern depend on the amount of energy available in an arbitrarily chosen region of the universe? This, to repeat, would be just as arbitrary as relativizing the reality of patterns to the computational capacity of humans. What we need is a way to

formulate the indispensability or non-redundancy of patterns that does not appeal to the notion of physical possibility. We will return to this issue in Chapter 3, where we introduce the concept of RP emergence.

§1.4 Multiple Instantiability

We now turn to an important characteristic of patterns which I believe has not received sufficient emphasis in the literature. This is what I shall call their *multiple instantiability*, that is, the capacity for the same pattern to be instantiated by different particular elements. In the introduction I mentioned the example of a melody played on the piano: the melody will retain its identity even if each of its individual notes is raised or lowered an octave. Our body is also multiply instantiable in this sense, since it remains the same even though the cells that compose it are constantly being replaced by new ones. I contend that any pattern one could think of has this property: the same flocking pattern can be exhibited by different particular birds; a wave propagating through a gas or liquid persists even as the molecules that constitute the wave constantly change; the solubility of salt in water manifests itself every time the relevant conditions are fulfilled; etc.

There are three points of clarification that I want to make here. First, a difference between the flocking pattern and the wave is that in any given instance of flocking behavior, the birds constituting the overall flocking pattern remain fixed, whereas in the case of a wave, the molecules constituting the wave pattern change over time. Humphreys (2008:437–38) refers to patterns of the first kind as *micro-stable patterns* and those of the second kind as *micro-dynamic patterns* (and further distinguishes three sub-types of the latter). Our body is also an example of a micro-dynamic pattern. This difference, however, does not make the flocking pattern any less multiply instantiable than the wave or body is.

Second, it may be argued that by saying that our body or a wave is multiply instantiable, I am confusing the instantiation relation with a part-whole relation. The cells that compose our

body, and the molecules that constitute a wave, are in a part-whole relation with the overall pattern, and the part-whole relation is not the same as the instantiation relation. My reply is that the constituent elements are in *both* a part-whole relation and an instantiation relation with the overall pattern. It seems undeniable that my body at a time t_1 and at a time t_2 a year later, say, are the same pattern instantiated at different points in time by different constituent elements; and the elements' being in a part-whole relation with the overall pattern does not prevent them from also being in a relation of instantiation with that same pattern.

Third, multiple instantiability should not be confused with *multiple realizability*, often discussed in connection with emergence. The difference between these two notions can be clarified using the type/token distinction, originally introduced by Peirce.[5] A pattern is multiply realizable if it can be instantiated by elements of different types. A given amount of money, for example, can be realized in the form of coins, bills, checks, or stored electronically in a bank account. In other words, the same amount of money can be realized by different *kinds* of physical entities. A pattern is multiply *instantiable*, on the other hand, if it can be instantiated by different *tokens* of some type of element (where the tokens need not be of the same type). Multiple realizability implies multiple instantiability, but not vice-versa. An example of a pattern that is multiply instantiable but not multiply realizable is a pattern generated in a cellular automaton with only one type of cell. The pattern is multiply instantiable, since the same pattern can be instantiated by different particular cells, but as long as there is only one type of cell, it is not multiply realizable.

We have thus established that patterns are multiply instantiable. But this is simply another way of saying that patterns are general, for the traditional definition of a general (or universal) ever since Aristotle is: a general is that which can be predicated of many things; or in other

[5] In "Prolegomena to an Apology for Pragmaticism" (CP 4.537, 1906). It is interesting to note that the type/token distinction was originally introduced as part of a trichotomy, type/token/tone, but *tone*, which corresponds to Firstness in Peirce's categorial scheme, was lost as the distinction entered the analytic philosophy literature.

words, a general is that which is multiply instantiable.[6] That patterns are general is not surprising if we consider the fact that to discern a pattern is to discern a certain *form*, and a form is general. Samuel Alexander, one of the fathers of British emergentism, explicitly draws a connection between patterns and generals, arguing that the "quality" that emerges in any instance of emergence is at once a pattern and a universal: "To adopt the ancient distinction of form and matter, the kind of existent from which a new quality emerges is the 'matter' which assumes a certain complexity of configuration and to this pattern or universal corresponds the new emergent quality" (Alexander 1920, 2:47). Given that patterns are generals, Dennett's theory of real patterns can be said to be a revival, in modern garbs, of Scholastic realism—the doctrine that there are real generals—although I doubt that Dennett himself views his theory in this way.[7] The generality of patterns makes them perfect candidates for applying Peirce's rich theory of generals, in particular his modal analysis of generality. To this we shall now turn.

[6] Aristotle, *De Interpretatione*, VII, 17a38.

[7] There is, however, a residue of nominalism in Dennett's theory, namely his retention of Hans Reichenbach's distinction between *illata* (concrete physical objects) and *abstracta* (abstract objects), and his characterization of the latter as "lossy compression[s]" (Dennett 2000:360). Ross (2000), rightly in my opinion, argues that Dennett should abandon the illata/abstracta distinction.

Chapter 2: Peirce's Theory of Generality

> *What is it that gives a certain sound and certain meaning to a mere jumble of lines? When the old scholar came to this thought, he acknowledged without hesitation the existence of letter-spirits. Just as a pair of hands and legs, a head, nails, and stomach not governed by a soul is not a human, so how could a mere collection of lines possess a sound and meaning, if it were not governed by a spirit?*
>
> Atsushi Nakajima, *Mojika* [The Curse of Letters]

This chapter will be devoted to a study of Peirce's theory of generality, as encapsulated in his pragmatism and Scholastic realism. We will begin by briefly laying out the terminology and basic framework of the problem of universals (§2.1). We will then turn to Peirce's unique brand of realism, focusing not only on how he argued for his own position, but also how he redefined the traditional problem of universals, as mentioned in the introduction. Our approach will be roughly chronological: we will begin by discerning hints of Peirce's realism in his earliest publication, "On a New List of Categories" (1867) (§2.2). We will then discuss his review of Alexander Campbell Fraser's (1819–1914) *The Works of George Berkeley* (1871), which is perhaps the clearest statement of his realism in his entire oeuvre (§2.3). Next we turn to Peirce's pragmatic maxim, and highlight the difference between what he calls *will-be's* and *would-be's* in his later works (§2.4). This distinction, as we shall see, constitutes an integral part of Peirce's modal analysis of generality. Finally, we will take up one of Peirce's late arguments for Scholastic realism, which has a striking similarity with Dennett's idea that the reality of patterns consists in their predictive power (§2.5). This argument will allow us to see why a pattern's being projectible into the future is a good indication of its being *real* in the sense of having a mind-independent being.

§2.1 The Problem of Universals

The dispute over universals revolves around a deceptively simple question: are there real generals?[8] A *general*, we may recall, is that which can be predicated of many things. *Humanity*, for example, can be predicated of Socrates, Plato, or any other human, and is therefore general. That which cannot be predicated of anything else, and can only be the subject of a predication, will be called a *particular*. Let us then consider what it means for something to be *real*. Peirce's definition, which he often attributes to the medieval philosopher-theologian John Duns Scotus (1265/66–1308),[9] is as follows: the real is "that whose characters are independent of what anybody may think them to be" (EP 1:137, W 3:271, 1878).[10] That is, something is said to be real if its being such as it is cannot be (or could not have been) altered by the mere act of thinking it to be otherwise. Thus, Prince Hamlet is not real, since his characters (that he is the Prince of Denmark, that he sees his father's ghost, etc.) could have been different if Shakespeare had conceived of him differently; but the very fact that Shakespeare wrote *Hamlet* is real, since nothing about that fact can be (or could have been) altered by the mere act of thinking it to be otherwise.

---

[8] The terms "general" and "universal" will be used interchangeably. "General" will be my preferred terminology, since this is the term favored by Peirce, but I will retain the term "universal" whenever it occurs in well-established phrases, such as "the problem of universals" or "the dispute over universals."

[9] In a 1909 manuscript entitled "Significs and Logic" (R 642:5), Peirce states that the closest Scotus comes to giving a definition of "reality" is in his *Opus Oxonienese*, Distinctio XIII, Quaestio iv, where Scotus writes: "Ens reale quod distinguitur contra ens rationis, est illud quod ex se habet esse circumscripto omni operae intellectus, ut intellectus est" (a real being, as contrasted with a being of reason, is that which has being of itself, setting aside all operations of the intellect insofar as it is intellect). In the Vatican edition of Scotus' works, in which the *Opus Oxoniense* is included as the *Ordinatio*, the passage quoted by Peirce occurs in *Ordinatio* I 8.177. See also ILS 104–5, fn.19.

[10] Note that this definition of reality is not a "clear" definition in the context of the paper in which it appears ("How to Make Our Ideas Clear"). The "clear" definition of reality involves Peirce's convergence view of truth, which will be discussed in §2.3 below.

Taking *humanity* as our example again, our question can thus be rephrased as follows: is there something that Socrates, Plato, and every other human really have in common, independently of how we conceive of them, by virtue of which we are able to apply the same concept or word *humanity* to them; or is it the case that *humanity* is nothing but a mental or verbal sign that we apply to a class of particulars which in themselves have nothing really in common? The first answer is that given by the *realist*, while the second answer is that given by the *nominalist*.

Both realism and nominalism come in various forms. Two major forms of realism are Platonic realism (*ante rem* realism) and moderate realism (*in re* realism). The Platonic realist holds that generals are entities that subsist independently of their particular instantiations, while the moderate realist holds that generals subsist only in their particular instantiations. Thus, if we suppose that all humans disappear one day from the world, the moderate realist would hold that *humanity* too will disappear, whereas the Platonic realist would hold that *humanity* would continue to subsist even if there were no particular instances of it. It should be noted that the Scholastic dispute over universals that took place in the early 14th century was a dispute between moderate realists and nominalists; no one during that period supported Platonic realism. As noted by Marilyn McCord Adams: "Fourteenth-century 'moderate' realists agreed that natures must be somehow common to particulars in reality, but Aristotle had convinced them that no one in his right mind could hold that the nature of a thing exists separated from it as the Platonic forms were supposed to do" (Adams 1982:411).

Nominalists may also differ depending on how they account for the fact that we habitually apply the same general sign to a class of particulars which in themselves have nothing really in common. Perhaps the most common account is that we do so on the basis of a perceived similarity among the particulars. The general sign is then regarded as the result of abstracting away those features of the particulars in which they differ, and retaining only the aspect in which they all agree. The classic account of this abstraction procedure is given by John Locke (1632–1704)

in his *An Essay Concerning Human Understanding* (1689). There, he suggests that children acquire the idea of *Man* by first attending to particular persons such as their nurse or mother, and then observing that there are many other things in the world that resemble those individuals. Children are thus led to

> … frame an *Idea*, which they find those many Particulars do partake in; and to that they give, with others, the name *Man*, for Example. And *thus they come to have a general Name*, and a general *Idea*. Wherein they make nothing new, but only leave out of the complex *Idea* they had of *Peter* and *James*, *Mary* and *Jane*, that which is peculiar to each, and retain only what is common to them all. (Locke [1689] 1975: III, iii, §7)

Note that on this view, the similarity among the particulars can only be a similarity in *how they are perceived*, since it is the assumption of the nominalist theory that in themselves particulars have nothing really in common. The nominalist is thus able to claim that a general is nothing more than a mental or verbal sign that we apply to a class of particulars constituted by our perception of the similarity of those particulars.

In discussions of the problem of universals, one often encounters a third position that is distinct from both realism and nominalism. This is *conceptualism*, which holds that generals exist only as concepts or ideas and have no extra-mental reality. However, in this dissertation I will avoid speaking of conceptualism, for the following reason. While it is often taken for granted that if generals are of a mental nature, then they cannot be real, this is a *non sequitur*. Just because something is of a mental nature does not prevent it from being real (in the sense defined above); and indeed, Peirce himself often claims that generals are *ideas* that are sometimes real: "*Realism*, in the proper sense of the word, sanctioned by the continual usage of nigh a thousand years, is the doctrine that *reality* and *idea* are not contrary, but that *ideas* are sometimes *real*" (R 860:8, 1893).[11] The crucial question is not whether generals are of a mental nature, but whether

[11] See also Peirce's critique of Karl Pearson in his review of the latter's *The Grammar of Science* (EP 2:62–63, 1901).

or not they can be real. Introducing conceptualism as a third position only serves to muddle the issue.

Finally, it is important to note that the problem of universals has traditionally been framed as a question about the reality of genera and species, common natures, or in more recent times, properties and relations. This is a reflection of the fact that the problem has its roots in linguistic considerations. Whenever we make a judgement or assertion we use general terms, not only as predicates but also in the form of common nouns. General terms are the warp and woof of language; without them we would not be able to say or perhaps even think anything. It is only natural, then, that the question should arise as to whether the general terms we use correspond with anything in reality; and this way of asking the question leads us to formulate it in terms of genera and species, common natures, or properties and relations, since these are what we expect general terms to represent.

However, the problem of universals need not be framed in this way. We may recall that a general, in its minimal sense, is anything that is *multiply instantiable*; and this does not entail that generals must be genera and species, common natures, or properties and relations. Embracing a broader conception of generals may provide us not only with novel arguments in support of their reality, but also a better understanding of what their reality consists in. And this is precisely what Peirce does. As we will see in §2.4, his pragmatic maxim dictates that to judge that a certain general is applicable to some object x is to judge that x is governed by a series of certain laws or regularities. The question of the reality of generals is thus recast as a question about the reality of the corresponding laws and regularities, which can be argued on empirical grounds. But before going into Peirce's pragmatic reformulation of the problem of universals, let us take a look at a few other aspects of his realism, starting with his earliest published work, "On a New List of Categories" (1867).

§2.2 Peirce's Realism in the "New List of Categories"

In the remaining sections of this chapter we will follow the development of Peirce's realism in a roughly chronological order. Max H. Fisch ([1967] 1986) has argued that Peirce was initially a nominalist, and gradually progressed towards realism.[12] While it is true that he uses the word "nominalism" in an approving way prior to the cognition series of 1868, and while there were perhaps some nominalistic elements in Peirce's early thought, Don D. Roberts (1970) has convincingly argued that there is not enough evidence to claim that he was a nominalist even initially. Indeed, I think we should trust Peirce's authority when he says that "never, during the thirty years [1863–1893] in which I have been writing on philosophical questions, have I failed in my allegiance to realistic opinions and to certain Scotistic ideas" (CP 6.605, 1893). Nonetheless, it is clear that there were significant developments in Peirce's thought towards a more and more thorough realism, as argued by Fisch ([1967] 1986). Let us turn to these developments.

We can discern an implicit commitment to realism in Peirce's 1867 paper "On a New List of Categories" (hereinafter simply "New List"), which—along with four other papers published in the same issue of the *Proceedings of the American Academy of Arts and Sciences*—is his earliest published work. Here we will not go into the argument of the "New List" in any detail. Instead, we will content ourselves with a few remarks on the general purpose of the paper, and also a broad outline of the theory of cognition presented in it. These brief comments will be confined to the extent necessary in understanding the realism implicit in this work; the reader is referred to the supplementary chapter of this dissertation for a more detailed exposition.[13]

Peirce's aim in the "New List" is to identify and derive what he calls the *categories*, and thereby explicate the structure of the cognitive process at its most fundamental level. Following

[12] By "initially" he means the time of Peirce's first professional publications in logic and philosophy, i.e. 1867. According to Fisch, this initial nominalist period lasted until 1868, when the three papers of the cognition series appeared in the *Journal of Speculative Philosophy*.

[13] See also Murphey (1965), Murphey (1993, Chap. III), Ransdell (1966), Buzzelli (1972), Michael (1980), De Tienne (1996), and Ishida (2009).

Immanuel Kant (1724–1804), he defines a category as a *universal conception* (EP 1:1, W 2:49, 1867), where "universal" means that the conception is operative in every act of cognition whatsoever.[14] He identifies five categories, arranged in a hierarchical order of increasing abstractness. In the order from the most abstract to the least abstract ("nearest to sense"), the categories are as follows (EP 1:6, W 2:54, 1867):

Being
    Quality (Reference to a Ground)
    Relation (Reference to a Correlate)
    Representation (Reference to an Interpretant)
Substance

In his later works Peirce will drop *being* and *substance* from the list of categories, and the three intermediate categories (in the order listed) become his famous triad of *Firstness*, *Secondness*, and *Thirdness*, respectively. Here I want to focus on §7 of the "New List," where Peirce discusses the first of the three intermediate categories, namely *quality* or *reference to a ground*.

According to the theory of cognition developed in the "New List," whenever we cognize something, the process of cognition begins with the presentation of an undifferentiated manifold of sense impressions, which Peirce calls "substance." This manifold, which is initially a confused collection of impressions, is ordered into a single unified experience through the application to it of various conceptions; Peirce refers to this process as the "reduction of the manifold to unity." The cognitive process—the reduction of the manifold to unity—ends with the formation of a proposition (or judgement). A proposition consists of a predicate conjoined to a subject, where the substance (manifold) plays the role of the subject. The proposition is formed by applying some quality as a predicate to the substance. Consider, for the example, the proposition "this stove is black," which Peirce discusses in §7 (EP 1:4, W 2:52–53). Here, the "this" corre-

[14] Peirce gives the following definition of a "universal" conception in Lecture IX of his 1866 Lowell Lectures: "Of the numerous conceptions of the mind, some apply only to certain special collections of impressions and are called *particular*. Others apply to all collections of impressions and are called *universal*" (W 1:473).

sponds to the substance, and the proposition is formed by applying two qualities in succession to this substance: the quality of *being a stove* and the quality *black*.[15]

Peirce refers to quality as "reference to a ground." A *ground* is the result of what Peirce will in his later works call *hypostatic abstraction*—the procedure of creating a new abstract entity by hypostatizing a property, relation, or operation. In his 1906 paper "Prolegomena to an Apology for Pragmaticism," Peirce describes hypostatic abstraction as a process of "turning predicates from being signs that we think or think *through*, into being subjects thought of" (CP 5.549). The transformation of "honey is sweet" into "honey possesses sweetness" is an example of hypostatic abstraction. In his later works Peirce calls the new entity created by this procedure a "hypostasis" (EP 2:394, 1906), but in the "New List" it is referred to as a "pure abstraction" (EP 1:4, W 2:52, 1867).

Let us return to the proposition "this stove black." The quality *black* is obtained by referring to the pure abstraction *blackness*, and in this sense *black* can be said to be a reference to *blackness*. *Blackness* is the ground, reference to which constitutes the quality of being *black*. A proposition asserts that a certain quality is applicable to the substance. In order for this to be *asserted*, the quality must be apprehended not as applied to a particular instance, but *in itself*, without regard to any specific circumstance (EP 1:4, W 2:52, 1867). A quality apprehended in itself, independently of any specific circumstance, is a pure abstraction like *blackness*.

Peirce makes the following important point in §7: "the conception of a pure abstraction is indispensible, because we cannot comprehend an agreement of two things, except as an agreement in some *respect*, and this respect is a pure abstraction as blackness" (EP 1:4, W 2:52, 1867). Although André De Tienne regards the view expressed in this passage as constituting "the major nominalist element of the otherwise realist epistemology of the young Peirce" (De

[15] Note that Peirce does not speak of successively applying two qualities in this context. Rather, he seems to assume that the quality of *being a stove* has already been applied to the substance, and focuses on the application of the quality *black*. I have added the extra step for the sake of clarity.

Tienne 1996:282), I see it as rather indicative of Peirce's anti-nominalist stance. As noted in the previous section, a typical nominalist strategy in accounting for our use of general signs is to try to reduce generals to perceived similarity relations among particulars. For example, consider the process by which we come to acquire the general sign *black*. We first perceive a similarity among particular black things. We group these things together into one class, and then apply the sign *black* to this class. Psychologically speaking, this may be a more or less accurate account of the way we learn general signs. The nominalist, however, draws from this the conclusion that generals have no reality—they are nothing but mental or verbal signs that we apply to certain classes of particulars. On the nominalist view, it is not the case that certain particulars resemble one another because they share a common attribute; rather, they are perceived to share a common attribute because they are perceived to resemble one another.

Against this view, Peirce is claiming that a resemblance can only be understood as an agreement in some *respect*, and this respect is a pure abstraction like *blackness*. The judgement that two or more things resemble each other should not be made in an arbitrary way—there should always be a *ground* for the judgement. We are able to judge that two or more things resemble one another only because we have prior access to a pure abstraction like *blackness*, and we recognize that this same pure abstraction is embodied (hypothetically) in each of the particular objects. The same can be said for the case in which we make a judgement about a single object. Hence Peirce's claim that reference to a ground is a universal conception (category), operative in every act of cognition whatsoever. Any judgement, insofar as it is not made in a completely arbitrary way, should be made on the basis of some ground, and this ground is a quality apprehended in itself, independently of its application to any particular instance; in other words, it is a pure abstraction. While Peirce does not explicitly raise the issue of the reality of generals in the "New List," the view expressed here can be said to be a realist one, insofar as it holds that there must be an abstract entity—a "pure abstraction"—that is in some sense prior to each of its concrete instantiations.

§2.3 The Berkeley Review

The first notable step in the development of Peirce's realism can be seen in his 1871 review of Fraser's *The Works of George Berkeley*. In this article, Peirce traces the history of the dispute over universals leading up to George Berkeley (1685–1753) in order to throw light on the latter's philosophy. However, not only does Peirce add an original twist to his account of the problem of universals; he also puts forward his own unique brand of realism. Indeed, although clothed as a book review, this article is perhaps the clearest statement of Peirce's realism in his entire oeuvre. As we will see, an essential component of Peirce's realism is a form of idealism which Robert Lane has called "basic idealism" (Lane 2018:60)—the view that anything real must be a possible or actual object of thought.

Peirce's account of the problem of universals in this article begins with a straightforward statement of the problem as it has been traditionally conceived:

> The question, therefore, is whether *man*, *horse*, and other names of natural classes, correspond with anything which all men, or all horses, really have in common, independent of our thought, or whether these classes are constituted simply by a likeness in the way in which our minds are affected by individual objects which have in themselves no resemblance or relationship whatsoever. (EP 1:88, W 2:467, 1871)

He then goes on to add his own twist to this traditional conception. He notes that both realists and nominalists agree that there must be such a thing as *reality*, something independent of how we think of it, for otherwise there would be nothing constraining our opinions, and all our thoughts would be arbitrary fictions. Where the two views differ is how they conceive of this reality. One view of reality, a "very familiar one" (EP 1:88, W 2:468, 1871), goes something like this:

> We have … nothing immediately present to us but thoughts. Those thoughts, however, have been caused by sensations, and those sensations are constrained by something out of the mind. This thing out of the mind, which directly influences sensation, and through sensation thought, because it *is* out of the mind, is independent of how we think it, and is, in short, the real. (EP 1:88, W 2:468, 1871)

According to Peirce, this view, which locates reality in the external objects that produce our sensations, is the nominalist conception of reality. It is nominalist because it entails that there can be no real generality. It is true that on this view, one might admit as a rough statement that two men are similar, the exact sense being that they produce sensations that we embrace under the same mental (or verbal) sign. However, insofar as this view conceives of similarity in this way, it cannot admit that there is something real that the two men have in common. As Peirce argues:

> [I]t can by no means be admitted that the two real men have really anything in common, for to say that they are both men is only to say that the one mental term or thought-sign "man" stands indifferently for either of the sensible objects caused by the two external realities; so that not even the two sensations have in themselves anything in common, and far less is it to be inferred that the external realities have. (EP 1:88, W 2:468, 1871)

One problem with a view like this is that it leads to skepticism about our knowledge of the world. We do not have cognitive access to the external objects that produce our sensations, and so there is no guarantee that our sensations "copy" or "represent" these objects exactly as they are. Hence our thoughts, which on this view are supposed to be constrained by our sensations, have no foothold on reality. By driving a stake in between thought on the one hand and reality on the other—by conceiving of reality as a *Ding an sich*—this view makes the latter something utterly unknowable.

This leads to a further, perhaps more serious problem. In his 1868 essay "Questions Concerning Certain Faculties Claimed for Man," Peirce had argued that to posit something absolutely incognizable is self-contradictory (EP 1:25, 1868). To think of something as being beyond all thought is to think it nonetheless, and so the conception of that thing must have the form "A, not-A." In order for something to have any being at all, it must be a possible object of thought. Even if it is unknown at a certain stage of inquiry, it must in principle be cognizable through inquiry. Thus Peirce concludes: "Over against any cognition, there is an unknown but knowable

reality; but over against all possible cognition, there is only the self-contradictory. In short, *cognizability* (in its widest sense) and *being* are not merely metaphysically the same, but are synonymous terms" (EP 1:25, W 2:208–9, 1868). This thesis constitutes Peirce's *basic idealism* that was mentioned at the beginning of this section. Notice the parallel with Dennett's observation that a pattern, by definition, must be a candidate for pattern *recognition*—an incognizable pattern is a contradiction in terms (§1.2).

What, then, is the realist conception of reality? First Peirce observes that there is an element of error in all human thought—an "arbitrary, accidental element, dependent on the limitations in circumstances, power, and bent of the individual" (EP 1:89, 1871). This gives the realist a clue as the where to locate the real. Whereas the nominalist seeks the real in the past, conceiving it as the cause of our sensations, the realist locates the real in the future, for it is in the future that we expect the errors in thought to be ultimately eliminated.

Experience shows us that although our initial impressions or opinions with respect to a given question may be affected to a large degree by our individual idiosyncrasies, the conclusion we reach after a thorough inquiry into the question will often coincide with that reached by other inquirers. Peirce illustrates this point using the example of a blind man and deaf man:

> Suppose two men, one deaf, the other blind. One hears a man declare he means to kill another, hears the report of the pistol, and hears the victim cry; the other sees the murder done. Their sensations are affected in the highest degree with their individual peculiarities. The first information that their sensations will give them, their first inferences, will be more nearly alike, but still different; the one having, for example, the idea of a man shouting, the other of a man with a threatening aspect; but their final conclusions, the thought the remotest from sense, will be identical and free from the one-sidedness of their idiosyncrasies. (EP 1:89, W 2:468–69, 1871)

From this Peirce argues that "to every question [there is] a true answer, a final conclusion, to which the opinion of every man is constantly gravitating" (EP 1:89, 1871). This final opinion, the ideal limit of inquiry, is what we call the *truth*, and what is represented in this opinion is what we call *reality*.

General agreement in the final opinion may be postponed, perhaps even indefinitely, due to "the arbitrary will or other individual peculiarities of a sufficiently large number of minds" (EP 1:89, 1871). This, however, cannot affect what the final opinion will be when it is reached. Peirce thus concludes that the final opinion "is independent, not indeed of thought in general, but of all that is arbitrary and individual in thought; is quite independent of how you, or I, or any number of men think" (EP 1:89, 1871). It is in this sense that the final opinion (and hence what is represented in it) can be said to be *real*. Notice Peirce's remark that the final opinion is not independent of "thought in general." The final opinion, insofar as it is an *opinion*, is of an intellectual nature, and so cannot be independent of all thought. This is consistent with Peirce's basic idealism, which holds that nothing can be independent of all thought. On the other hand, this does not prevent the final opinion (and what is represented in it) from being real in the sense described above.

Peirce further notes that this conception of reality entails that there are real generals, because "general conceptions enter into all judgments, and therefore into true opinions" (EP 1:90, 1871). For example, the statement "all white things have whiteness in them" is true—we expect it to hold in the final opinion—because this is simply another way of saying "all white things are white." And since it is true that there are real things that possess whiteness, whiteness is real (EP 1:90, 1871).

While Peirce attributes the view of truth and reality articulated above—the former of which is often referred to as the *convergence theory of truth* in the philosophical literature—to the medieval realists, no medieval philosopher explicitly held such a view. The convergence view of truth and reality outlined in the Berkeley review should be regarded as a genuinely Peircean contribution. Nonetheless, Peirce's observation that nominalists and realists are divided by a fundamental difference in their understanding of the notion of reality, I believe, captures the often implicit line of thought at the heart of the dispute over universals.

An objection against Peirce's convergence view of truth and reality that readily suggests itself is the following: on what basis is he able to claim that that "to every question [there is] a true an-

swer, a final conclusion, to which the opinion of every man is constantly gravitating"? How do we know that there is a true answer to every question? Furthermore, if truth is to be identified with the final opinion, then how are we to make sense of "buried secrets" that we presumably have no way of finding out, such as the exact number of hairs that were on the head of Julius Caesar when he crossed the Rubicon? Common sense tells us that there must be a true answer to the question, "how many hairs were on Julius Caesar's head when he crossed the Rubicon?" Isn't this an instance of a truth that no amount of inquiry can ever uncover?

Peirce himself seems to have come to see the unwarranted optimism of his above claim, for in his later writings he abandons this view, arguing instead that it is a regulative assumption of inquiry, an "intellectual hope" (EP 1:275, 1887–88), that to any given question there is a true answer which can be discovered if inquiry into that question were pursued far enough. There is no guarantee that this is actually so. There may be meaningful questions that no amount of inquiry can ever settle: "there is not the smallest scintilla of logical justification for any assertion that a given sort of result will, as a matter of fact, either *always* or *never* come to pass; and consequently we cannot know that there *is* any truth concerning any given question" (EP 2:419, 1907). However, whenever we inquire into a question, we proceed upon the *assumption* that it has a definite answer, and that this answer can be attained sooner or later through inquiry, for to suppose otherwise would be a self-stultification.[16]

There are several other objections that have been levelled against Peirce's convergence theory of truth and reality, but this is not the place to go into these objections.[17] What I want to do instead is highlight the proximity of Peirce's brand of Scholastic realism and Dennett's theory of real pat-

[16] For more detailed discussions of Peirce's notion of a regulative assumption and his reply to the "buried secrets" objection, see Misak (2004:67–70, 137–42) and Lane (2018:51–58, 165–69).

[17] W. V. O. Quine, for example, criticizes Peirce's convergence theory of truth in *Word and Object*, arguing that "[t]here is a faulty use of numerical analogy in speaking of a limit of theories," and also that "there is trouble in the imputation of uniqueness" to the ideal limit of inquiry (Quine 2013:21). See Ishida (2013) for a Peircean reply to Quine's objections.

terns. As we saw above, the crucial question that separates the nominalist and realist, according to Peirce, is whether one should admit a *Ding an sich*, an external reality independent of all thought. The nominalist answers this question in the affirmative, whereas the realist answers it in the negative, seeking instead to define reality solely in terms of what can in principle be cognized. The same can be said of Dennett's theory of patterns: recall that he defines the reality of a pattern in terms of whether discerning it will lead to successful predictions, which is something perfectly cognizable, and *not* in terms of whether it faithfully represents or corresponds with some deeper but inaccessible reality.[18] Here then is another aspect in which Dennett's theory of real patterns can be said to be a modernized version of Scholastic realism (as understood by Peirce).

So far we have looked at some of the facets of Peirce's early realism, leading up to his 1871 Berkeley review. In the next section I want to turn to Peirce's pragmatic maxim and his modal analysis of generality, which will play a central role in our formulation of RP emergence in the following chapter.

§2.4 The Pragmatic Maxim

In §1.2 we saw that according to Dennett's criterion, a real pattern is one whose discernment allows us to make successful predictions about future events. For Peirce, however, real generals (including real patterns) have a richer content than this: they not only support predictions about what *will* happen in the future; they also give us information about what *would* happen in conceivable circumstances that are not actualized. Or as Peirce would say, a general is not merely a *will-be* but a *would-be*. As I hope to show in the following chapter, Peirce's modal analysis of generality, according to which generals have the modality of *would-be's*, will give us a clear understanding of the autonomy characteristic of RP emergence.

[18] In fact, in the particular context of discussing the reality of beliefs and other intentional states, Dennett explicitly argues against such a view, which he labels "industrial-strength Realism" and attributes to Jerry Fodor (Dennett 1991:42–45).

Peirce's modal analysis of generality is intimately connected with his pragmatism, so we shall begin by considering his pragmatism. Peirce himself was not always clear on the difference between *will-be's* and *would-be's*. In his early years he tended to vacillate between using the indicative mood (*will be*) and subjunctive mood (*would be*) in stating pragmatic clarifications.[19] It was only in his later years, in the 1900s, that he became explicit about the difference between the two and began stressing the importance of using the subjunctive mood in stating pragmatic clarifications. A good way to understand the difference between *will-be's* and *would-be's*, therefore, is to trace the development of Peirce's own ideas on the subject.

Peirce's pragmatic maxim appeared in public form for the first time in his 1878 paper "How to Make Our Ideas Clear" (HTM), published as part of the *Illustrations of the Logic of Science* series in the *Popular Science Monthly*. It was formulated as a logical principle for clarifying the meaning of ideas, for attaining the "third grade of clearness" of apprehension, the first two being the traditional criteria of *clearness* and *distinctness* as formulated by Descartes and developed by Leibniz (EP 1:124–27, W 3:257–61, 1878).[20] The famous statement of the maxim runs as follows: "Consider what effects, which might conceivably have practical bearings, we con-

[19] See Lane (2018:51–58) for an account of this vacillation.

[20] Note that for Peirce, pragmatism is not a theory of *truth*, as it is in the case of William James (1842–1910) and John Dewey (1859–1952). James, for example, in a series of lectures published as *Pragmatism: A New Name for Some Old Ways of Thinking* (1907), writes: "*ideas (which themselves are but parts of our experience) become true just in so far as they help us to get into satisfactory relation with other parts of our experience*, to summarize them and get about among them by conceptual short-cuts instead of following the interminable succession of particular phenomena. Any idea upon which we can ride, so to speak; any idea that will carry us prosperously from any one part of our experience to any other part, linking things satisfactorily, working securely, simplifying, saving labor; is true for just so much, true in so far forth, true *instrumentally*" (James 1987:512, emphasis in original). This kind of "instrumental" view of truth is what people often seem to have in mind when they speak of *pragmatism*. It is important to keep in mind, however, that Peirce's pragmatism is a theory of *meaning* rather than a theory of truth—though his view of truth as the ideal limit of inquiry that we discussed in the previous section may be regarded as the result of applying the pragmatic maxim to the concept of truth.

ceive the object of our conception to have. Then, our conception of these effects is the whole of our conception of the object" (EP 1:132, W 3:266, 1878).

To use the well-known example that Peirce himself gives, the meaning of the concept of *hard* is that an object to which *hard* can be applied as a predicate, such as a diamond, would not be scratched even if we apply pressure to it with, say, a knife-edge. That is, the meaning of the concept of *hard* can be analyzed into general laws expressible in conditional propositions of the form:

> If you were to apply pressure to *X* (where *X* is an object to which *hard* can be veritably applied as a predicate), then *X* would resist the pressure.

Notice that I have formulated this conditional in the subjunctive mood (*would*) rather than the indicative mood (*will*); this is in accordance with Peirce's later, more considered view. Now let us consider the general case. The pragmatic maxim states that the meaning of a concept can be analyzed into general laws expressible in conditional propositions of the form:

> (∗) If you were to do *m* to *X* (where *X* is an object to which the concept in question can be veritably applied as a predicate), then you would have an experience of type *n*.

We perform a certain kind of operation on nature (for example, we apply pressure to a diamond), and nature gives us a certain kind of response (the diamond resists the pressure). The *general law* that operations of a certain kind are always (or with a certain degree of probability) followed by responses of a certain kind is what constitutes the meaning of the concept in question.

What, then, is the difference between formulating pragmatic clarifications in the indicative mood and in the subjunctive mood? Indicative-mood clarifications analyze the meaning of a concept into laws that dictate what *will* happen in the future; whereas subjunctive-mood clarifi-

cations analyze the meaning of a concept into laws that have reference not only to what will happen in the future, but also to what *would* happen in possible situations that are not actualized. This difference is well-illustrated by a thought experiment that Peirce poses in HTM. There, he supposes that a diamond is burned up before its hardness could be tested, and asks whether it would be false to say that the diamond was soft. His answer, at the time of writing HTM, was "no":

> [T]here would be no *falsity* in such modes of speech. They would involve a modification of our present usage of speech with regard to the words hard and soft, but not of their meanings. For they represent no fact to be different from what it is; only they involve arrangements of facts which would be exceedingly maladroit. This leads us to remark that the question of what would occur under circumstances which do not actually arise is not a question of fact, but only of the most perspicuous arrangement of them. (EP 1:132, W 3:267, 1878)

In his 1905 paper "Issues of Pragmaticism," however, Peirce admits that this conclusion was a mistake, and argues that a diamond's hardness is a real property, regardless of whether it is actually put to the test or not (EP 2:356–57, 1905). The crucial difference between Peirce in the 1870s and Peirce in the 1900s lies in his conception of the kind of modality that general laws possess. For the earlier Peirce general laws are *will-be's*, that is, they have reference only to events that we know will occur in the future, while for the later Peirce general laws are *would-be's*, that is, they have reference also to counterfactual situations.

Peirce's conception of modality in the 1870s also appears in a subtle way in his 1871 Berkeley review. There, after having formulated his convergence view of truth and reality, he goes into a discussion of the concept of *power*:

> What is the POWER of external things, to affect the senses? To say that people sleep after taking opium because it has a soporific *power*, is that to say anything in the world but that people sleep after taking opium because they sleep after taking opium? To assert the existence of a power or potency, is it to assert the existence of anything actual? Or to say that a thing has a potential existence, is it to say that it has an actual existence? In other words, is the present existence of a power anything in the world but a regularity in future events relating to a certain thing regarded as an element which is to be taken account of beforehand, in the conception of that thing? If not, to

> assert that there are external things which can be known only as exerting a power on our sense, is nothing different from asserting that there is a general *drift* in the history of human thought which will lead it to one general agreement, one catholic consent. (EP 1:89–90, W 2:469, 1871)

The example of opium is one that Peirce takes up frequently in his writings. What is noteworthy here is that his answer to the question "To say that people sleep after taking opium because it has a soporific power, is that to say anything in the world but that people sleep after taking opium because they sleep after taking opium?" is "no," that is, soporific power is *nothing but* a certain regularity in future events, in this case that people sleep. But a power is a general law, and to say that it is nothing but a regularity in future events is, from the standpoint of Peirce's mature view, to reduce laws to *will-be's*.[21]

While the pragmatic maxim was originally formulated as a rule for clarifying the meaning of *concepts*, it can be extended into a principle for identifying the intellectual purport of any kind of general, anything capable of functioning as a predicate, including patterns. Together with Peirce's mature conception of general laws as *would-be's*, this implies that the intellectual purport of any general, and hence any pattern, is a *would-be*.[22] Thus, to judge that something has a certain general form, or that something exhibits a certain pattern, is to judge that it is governed by general laws expressible in subjunctive conditionals of the form $(*)$. From this it further follows that to discern a pattern in some system is to make predictions about how the system would behave under certain conceivable conditions. As we saw in the previous chapter, this is

[21] Compare this with Peirce's late views on the example of opium, for example in CP 4.234 (1902), PM 71–72 (1903), CP 5.534 (1905), and EP 2:394 (1906).

[22] It is not quite clear whether Peirce himself intended the pragmatic maxim to be a rule about the intellectual purport of any general. In a later formulation, the maxim is framed as a rule for identifying the "intellectual purport of any symbol" (EP 2:346, 1905), but not all generals are symbols, since in Peirce's famous ten-fold classification of signs, there are legisigns (general signs) that are not symbols (EP 2:289–99, 1903). On the other hand, legisigns are general laws that are signs, and hence have the modality of *would-be's*. Therefore, assuming that all generals are legisigns, it seems that for (the later) Peirce all generals have the modality of *would-be's*.

precisely what Dennett's theory of patterns asserts, except that here the predictions have reference not only to future events, but also to counterfactual situations. The similarity with Dennett's theory becomes even more striking when we consider one of Peirce's late arguments for Scholastic realism, which is the topic of the following section.

§2.5 Peirce's Late Argument for Scholastic Realism

So far we have seen that according to the pragmatic maxim, to predicate a general of some object is to judge that the object is under the governance of certain *would-be's*. Not all *would-be's*, however, are *real* (in the Peircean sense of the term; see §2.1). Recall the example of throwing a die that we considered in §1.2. To say that the die is *unbiased* is to ascribe to the die a real *would-be*, namely, that if it were thrown many times, each face would turn up with a relative frequency close to 1/6, the deviation from 1/6 becoming smaller and smaller as more throws are made. But if we were to say, after having observed only the first ten throws, that the die is *biased* in such a way that it only turns up sixes, then the *would-be* we ascribe to the die—that it would turn up a six every time it is thrown—would not be a real *would-be*, but a hasty generalization based on insufficient data.

This difference can also be expressed by saying that the property of being *unbiased* is a real property of the die, whereas the property of being *biased* is not. The pragmatic maxim analyzes the intellectual purport of any general predicated of a given object into a set of *would-be's*, but it does not by itself determine whether the *would-be's* are real or not; nor does it determine whether the general is a real property of the object or not. We therefore need some other criterion by which we can determine the reality of a general predicated of a given object. This criterion is provided by one of Peirce's late arguments for Scholastic realism. As was mentioned in §1.2, this argument is intimately related to Dennett's formulation of the reality of patterns in terms of predictive power. In fact, we will see that a pattern's being projectible into the future is a good indication of its being real in Peirce's sense.

The argument I want to take up appears in the "Seven Systems of Metaphysics," the fourth of Peirce's 1903 Harvard Lectures on pragmatism. There, he takes a stone in his hand and announces to the audience that he will perform an experiment: he will let go of the stone and see whether it will fall to the floor (EP 2:181, 1903). The experiment, of course, is meaningless, since everybody knows what will happen. But the deeper meaning of the experiment lies in the very fact that it is meaningless. How is it that we do not have to actually perform the experiment in order to know its result? The answer must be that the stone is governed by a real law operative in nature. If the law were only a mental or verbal formula and not real, there would be no way of explaining why future events will conform to it (and we know that they will), unless we were to suppose that the mind had some kind of miraculous power of prognosis. If, on the other hand, we suppose that the stone is governed by a real law dictating what would happen in certain kinds of situations, then witnessing the actual instantiations of the law will be no wonder. Now since laws are generals, it follows that there are real generals. Thus runs Peirce's argument for realism from our experience of anticipation.

Now according to the (extended) pragmatic maxim, the intellectual purport of any general consists in laws expressible in subjunctive conditionals of the form $(*)$. In the case of a stone, judging that something is a stone involves identifying the laws that it conforms to, one of which can be expressed as: "if you were to let go of the stone, it would fall," and the conjunction of all such laws constitutes the entire intellectual purport of the general form or pattern of *stoneness*. Therefore, insofar as we know those laws to be real, then so is the general form or pattern whose intellectual purport these laws constitute. We are thus able to see why a pattern's being projectible into the future is a good indication (but not conclusive proof—there is no such thing as a conclusive proof in matters empirical) of its being real in Peirce's sense. It should be noted that although the predictions associated with a given pattern have reference to counterfactual situations, the reality of the pattern can be assessed only by testing whether the predictions it affords

are actually fulfilled. It is only after we have established the reality of the pattern that we are justified in further generalizing the predictions to counterfactual situations.

Chapter 3: Real Pattern Emergence

> *Get rid, thoughtful Reader, of the Ockhamistic prejudice of political partizenship that in thought, in being, and in development the indefinite is due to a degeneration from a primal state of perfect definiteness.*
>
> Charles S. Peirce, "Some Amazing Mazes, Fourth Curiosity"

In this chapter we will begin our exploration of the concept of emergence. My aim here is to outline the notion of RP emergence using the ideas developed so far in this dissertation. In §3.1 I will discuss the basic features of what are commonly referred to as emergent phenomena, introduce the distinction between *epistemological emergence* and *ontological emergence*, and explain why the views that I call *ontological reductionism* and *radical emergentism* are both problematic. Next, I will set forth some preliminary observations on the notion of "levels," often presupposed in discussions of reduction and emergence, and offer my reasons for avoiding talk of levels (§3.2). Then I will present a formulation of RP emergence, and show why it is distinct from both epistemological and ontological emergence (§3.3). Finally, in §3.4, I will try to throw what I call RP emergence into sharper relief by comparing it with Mark Bedau's related notion of *weak emergence* (Bedau 1997, 2002).

§3.1 Epistemological and Ontological Emergence

The notion of emergence typically arises when we are dealing with physical or computational systems consisting of multiple elements, which may (but need not) be mutually interacting. In the broadest and barest sense of the term, an *emergent phenomenon* can be characterized as any global property or behavior of a multi-element system that is not exhibited by any of the constituent elements of the system in isolation. Any situation in which such a phenomenon arises is said to be an instance of *emergence*. Thus, properties of liquid water such as transparency, viscosity, and surface tension are emergent phenomena, since the individual molecules that com-

pose water do not exhibit these properties. Some other major examples of emergence that are often discussed in the literature are listed below:

- Emergence of thermodynamic quantities such as temperature, pressure, and entropy
- Phase transitions and critical phenomena in ferromagnets and fluids
- Self-organizing patterns in systems far from thermal equilibrium (e.g. convection cells, oscillating chemical reactions, life)
- Emergence of complex patterns in cellular automata
- Emergence of "mental properties" from neural processes
- Emergence of social behavior patterns from interactions between individual humans (e.g. crowd behavior, market economy, internet)

Clearly, emergent phenomena are ubiquitous in the sciences as well as everyday life. Yet the concept of emergence is hard to pin down exactly, and has been the subject of vigorous debate among both philosophers and scientists. A feature often attributed to emergent phenomena is that they are in some sense "novel" or irreducible to their underlying elements. Two kinds of emergence are often distinguished in the literature according to how this irreducibility is construed: *epistemological emergence* and *ontological emergence*.[23] Epistemological emergence is a kind of emergence that is only in the eyes of the beholder. An epistemologically emergent phenomenon is one that can in principle be reduced to—predicted or derived from—its underlying elements, but is in practice irreducible to these elements due to epistemic limitations on the

[23] Of course, different authors prefer different terminology. What I am here calling "epistemological emergence" has also been called "weak emergence" (Chalmers 2006; not to be confused with Mark Bedau's version of weak emergence, which will be discussed in §3.4), "conservative emergence" (Seager 2012), and "conceptual emergence" (Humphreys 2016), while what I am here calling "ontological emergence" has also been called "strong emergence" (Chalmers 2006), "radical emergence" (Seager 2012), and "brute emergence" (Strawson 2006). It should be noted that Chalmers' notion of weak emergence is broader than what I call epistemological emergence, since he defines weak emergence in such a way that cases of strong emergence are also cases of weak emergence, whereas I prefer to define epistemological and ontological emergence to be mutually exclusive.

part of the observer (limitations in computational resources, knowledge about initial conditions, etc.). An ontologically emergent phenomenon, on the other hand, is one that is irreducible to its underlying elements even in principle.

Here, a remark on the phrase "in principle" is in order. While philosophers frequently employ this phrase without much comment, this notion should be handled with care. What does it mean to say that something is possible "in principle"? In the specific case of reduction, I suggest that a global property or behavior of a system is "in principle" reducible to its underlying elements if there is an effective procedure for deriving it from a description of the micro state of the system, or equivalently, if there is a computable function that takes the micro state of the system as input and returns the global state as output. Note that even if a macro phenomenon is established as "in principle" irreducible in this sense, there may nonetheless be a hypercomputer that can derive it from its underlying elements.[24]

Epistemological emergence seems to be the conception of emergence most widely accepted among philosophers and scientists, whereas it remains controversial whether there are any instances of ontological emergence in the actual world. Suppose there is no ontological emergence, and that every instance of emergence in the world is epistemological. Combined with the assumption that every emergent feature in the world ultimately emerges from the entities and laws of fundamental physics, this implies that every emergent feature in the world can in principle be reduced to the entities and laws of fundamental physics, even if reference to emergent phenomena is indispensable in practice. Let us call this view *ontological reductionism*. Although widely endorsed by philosophers and scientists, I find this view problematic.[25] Let me explain why by citing a thought experiment put forth by Dennett.

In his essay "True Believers," Dennett asks us to imagine that beings of vastly superior intelligence, say Martians, descend upon us (Dennett 1987:25). Suppose, he says, that they are "La-

[24] On hypercomputation, see Syropoulos (2008).

[25] The view is held, for example, by Anderson (1972), Weinberg (1987), and Seager (2012).

placean super-physicists, capable of comprehending the activity on Wall Street, for instance, at the microphysical level. Where we see brokers and buildings and sell orders and bids, they see vast congeries of subatomic particles milling about" (Dennett 1987:25). According to the ontological reductionist, these Martians would know everything there is to know about the world. However, Dennett points out that even if the Martians were able to comprehend and accurately predict everything that happens on Wall Street using their Laplacean methods, they would be missing something perfectly real if they did not also see us as intentional beings, that is, if they did not also see the *patterns* in human behavior that we describe in intentional terms, such as *believing that p* or *desiring q*. As he puts it:

> Take a particular instance in which the Martians observe a stockbroker deciding to place an order for 500 shares of General Motors. They predict the exact motions of his fingers as he dials the phone and the exact vibrations of his vocal cords as he intones his order. But if the Martians do not see that indefinitely many *different* patterns of finger motions and vocal cord vibrations—even the motions of indefinitely many different individuals—could have been substituted for the actual particulars without perturbing the subsequent operation of the market, then they have failed to see a real pattern in the world they are observing. (Dennett 1987:26)

Although Dennett's argument is couched in terms of intentional patterns, the same argument can also be made with respect to any pattern outside the purview of fundamental physics—the Martians would be just as blind to fingers and vocal cords as they are to the intentional behavior of humans. This is a powerful thought experiment that shows us the inadequacy of ontological reductionism. It is inadequate because it fails to recognize that there is more to reality than what can be described at the level of fundamental physics. The patterns that we observe at everyday scales, as well as those studied by the various special sciences, are not mere epistemic crutches that can be dispensed with by hypothetical Laplacean super-physicists: they are just as real and fundamental as the patterns studied by fundamental physics.

Now if we are to deny ontological reductionism, it seems we ought to embrace the existence of ontological emergence.[26] However, the view that there are instances of ontological emergence in the world—which I will hereinafter call *radical emergentism*—is also problematic. The problem is that an ontologically emergent phenomenon is by definition utterly inexplicable, in the sense that there is absolutely nothing about the underlying elements by virtue of which it should emerge, and should have the features that it has. It is, in a word, sheer magic: it simply pops into existence without any why or wherefore. But to use a Peircean turn of phrase, to posit something utterly inexplicable is to set up a roadblock to inquiry. The synechistic philosophy demands that we do not introduce such brute discontinuities into the fabric of being.

Galen Strawson has gone further and argued that the notion of ontological emergence (which he calls "brute emergence") is incoherent:

> If it is really true that Y is emergent from X then it must be the case that Y is in some sense wholly dependent on X and X alone, so that all features of Y trace intelligibly back to X (where 'intelligible' is a metaphysical rather than an epistemic notion). *Emergence can't be brute*. It is built into the heart of the notion of emergence that emergence cannot be brute in the sense of there being absolutely no reason in the nature of things why the emerging thing is as it is (so that it is unintelligible even to God). For any feature Y of anything that is correctly considered to be emergent from X, there must be something about X and X alone in virtue of which Y emerges, and which is sufficient for Y. (Strawson 2006:18)

Perhaps one could take issue with Strawson's characterization of emergence, and define emergence in such a way as to make the notion of ontological emergence coherent. But even so, one cannot get around the fact that to posit an emergent phenomenon for which there is absolutely no reason in the nature of things why it is as it is—or in other words, to declare that a given emergent phenomenon can *never* be explained in terms of its underlying elements—is tantamount to abandoning inquiry into that phenomenon altogether.

[26] One could also attempt to deny the assumption, mentioned earlier, that every emergent feature in the world ultimately emerges from the entities and laws of fundamental physics. However, I will simply accept this as a plausible assumption.

If both ontological reductionism and radical emergentism are untenable, then it seems we are at an impasse. I suggest that the problem lies in the assumption that epistemological emergence and ontological emergence exhaust all conceivable forms of emergence. If there is a further, third form of emergence, then denying the existence of ontological emergence will not entail that every instance of emergence is epistemological. My task in this chapter will be to outline such a third form of emergence, inspired by the philosophical ideas of Peirce and Dennett, that will enable us to steer a path between the Scylla of ontological reductionism and the Charybdis of radical emergentism. Since what emerges in this form of emergence is a real pattern, I will simply call this form of emergence *real pattern emergence*, or *RP emergence* for short. But before we go into our discussion of RP emergence, a few remarks on the notion of "levels" are in order.

§3.2 The Notion of Levels

The notion that nature is organized in a hierarchical structure of "levels" is deeply embedded in discussions of emergence. One often speaks, for example, of "higher-level" entities possessing properties lacked by "lower-level" entities. However, it is by no means evident what levels are—are they objective features of nature, or do they somehow reflect the way we choose to describe nature?—or whether there are such things as levels at all. Indeed, Ladyman and Ross have denied the existence of levels, arguing that talk of "levels" is a metaphor to which contemporary science gives no interesting content (Ladyman & Ross 2007:53–57). Despite the elusiveness of the levels concept, most discussions of emergence in the philosophical literature simply assume that there are such things as levels, without addressing the issue of what they are and whether they actually exist.[27] It therefore behooves us to undertake a preliminary examination of the notion of levels before delving into our discussion of emergence.

[27] Some rare exceptions, in addition to Ladyman & Ross (2007), are Wimsatt (1976:237–263), Wimsatt (1994), and Humphreys (2016:120–26).

While I share Ladyman and Ross's skepticism regarding the notion of levels, it seems to me that their denial of the existence of levels needs to be qualified in at least two respects. In the first place, the relative strength and range of the four fundamental forces (the strong force, electromagnetic force, weak force, and gravity), together with the kinds of matter upon which they act, give rise to a separation of three "natural" levels or regimes, which I shall call the *subatomic regime*, *electromagnetic regime*, and *gravitational regime*.[28] Both the electromagnetic and gravitational forces have an infinite range and act on all size scales, but the electromagnetic force only acts on electrically charged matter, while gravity acts on every object in spacetime. Furthermore, in large objects the positive and negative electric charges tend to cancel each other out, making the object as a whole electrically neutral. This is why the influence of gravity tends to dominate at large scales, such as the scale of stars and galaxies. On the other hand, the electromagnetic force is much stronger than gravity, and so it tends to dominate at smaller scales, such as the scale of atoms, molecules, and what we regard as medium-sized objects. This is how the separation between the electromagnetic regime, dominated by the electromagnetic force, and the gravitational regime, dominated by gravity, arises. The boundary between these two regimes is by no means sharp, as is evinced by the fact that we experience the effects of both the electromagnetic force and gravity at everyday scales. The separation of the subatomic and electromagnetic regimes likewise arises from the fact that the influence of the weak and strong forces is confined to very small distances (of the order of 10^{-15} m, roughly the size of an atomic nucleus).

Thus, contrary to Ladyman and Ross's claim that levels do not exist, there are at least three levels or regimes in nature corresponding to three different size scales. However, levels identified solely in terms of size do not provide a sufficient basis for discussions of emergence. Some

[28] The discussion in this paragraph owes much to Reiji Sugano's study of the hierarchical structure of nature (Sugano 2013, Chapter 3). I also want to thank Taksu Cheon for his insightful suggestions on the topic in personal correspondence.

examples of emergence often discussed in the literature are the emergence of life from chemical processes, the emergence of mental phenomena from neural processes, and the emergence of social behavior patterns (such as crowd behavior) from interactions between individual humans. The problem is that there are no size scales that uniquely characterize organisms, mental phenomena, or social behavior patterns, and so if these are to be identified as levels, this identification cannot be made solely in terms of size. As pointed out by William Wimsatt (1994:236), a bacterium could have the same size as a black hole, but we would hardly consider the two as belonging to the same level, as they would behave in radically different ways in similar circumstances. As for mental phenomena and social behavior patterns, it is not even clear whether these could have sizes at all.

Evidently, size cannot be the sole factor in terms of which we identify levels. But there are other factors that enable us to do so, and this brings us to the second respect in which Ladyman and Ross's rejection of levels needs to be qualified. Consider, for example, the level of individual organisms. How do we identify this as a distinct level? I suggest that it is by focusing on a cluster of recurrent *patterns* or regularities that we observe in nature, such as reproduction, metabolism, and homeostasis. The level of individual organisms can be regarded as "higher" than the level of chemical processes because these patterns are lacking at the chemical level. In general, we can say that identifying a level involves picking out a set of patterns from the phenomena we observe in nature, and that a given level *A* is "higher" than another level *B* if the patterns in terms of which we identify *A* are lacking in *B*. This implies that what we identify as levels (apart from the three regimes mentioned above) depends to some extent on what patterns we detect and choose to focus on. Ladyman and Ross's rejection of levels has thus been qualified in two respects: in the first place, there are at least three regimes in nature corresponding to three different size scales; and in the second place, while it may be true that levels other than these three regimes do not exist apart from pattern-detecting agents, we can nonetheless speak intelligibly of these levels as long as we keep in mind their observer-dependent character.

Despite all of this, in this paper I will avoid talk of levels and speak instead of emergent patterns and their underlying elements and processes. The main reason for this is that the language of levels tends to give the impression that there is a fixed hierarchy of levels that somehow exists independently of the act of detecting and picking out the patterns that characterize them. This, however, is not the case: as I argued above, it is the act of detecting and picking out a certain set of patterns that gives rise to a distinct level; the level does not exist independently of this act. This is closely related to a point made by Paul Humphreys. A basic distinction between types of emergence (orthogonal to the epistemological/ontological distinction) is that between *synchronic emergence* and *diachronic emergence* (Humphreys 2008, 2016). In synchronic emergence, the emergent phenomenon is considered to exist simultaneously with its substrate, as when mental phenomena are considered to emerge from neural processes. In diachronic emergence, on the other hand, the emergent phenomenon is considered to develop over time from prior states of a system, as when complex patterns are generated in cellular automata. Humphreys argues that contemporary discussions of emergence in the philosophical literature have been overly focused on synchronic emergence (perhaps due to the circumstance that most philosophical discussions of emergence have taken place in the context of the philosophy of mind) and have neglected diachronic forms of emergence (Humphreys 2016). The notion of levels is one manifestation of this overemphasis of synchronic emergence, for as Humphreys points out, "the levels imagery is shot through with synchronic concepts" (Humphreys 2016:121). The levels framework is ill-suited to dealing with diachronic emergence. Again, the problem is that the language of levels tends to suggest that the "higher" level exists independently of the act of detecting and picking out the patterns that characterize it, whereas in diachronic emergence, where the emergent patterns come into being over time, it hardly makes any sense to say that there is a "higher level" when the patterns that allow us to identify it have not yet emerged. For these reasons I believe that the levels concept tends to distort rather than facili-

tate our understanding of emergence, and so I will avoid speaking of levels in this dissertation, except when presenting the views of other authors who use the concept.

With these preliminaries in place, let us turn to our main topic, RP emergence.

§3.3 Real Pattern Emergence

Employing the ideas developed in the preceding chapters, real pattern emergence may be defined as follows.

> A pattern *P* of a multi-element system *S* is said to be an *emergent real pattern* if it satisfies the following three conditions:
>
> (1) *P* is an emergent pattern of *S*, that is, *P* is a global pattern of *S* that is not exhibited by any of the constituent elements of *S* in isolation;
>
> (2) *P* can in principle be predicted or derived from knowledge about the constituent elements of *S*;
>
> (3) *P* is real in the sense that it supports predictions about not only what *will* happen given a certain state of *S*, but also what *would* happen in an indefinite variety of possible micro situations of *S* (i.e., possible states and arrangements of the constituent elements of *S*).
>
> Any instance in which a real pattern emerges in the above sense is an instance of *real pattern emergence* (*RP emergence*).

Condition (1) is simply a statement of the traditional notion of an emergent phenomenon, except that here it is framed in terms of patterns. By condition (2), RP emergence is similar to epistemological emergence and differs from ontological emergence, in that the emergent pattern can in principle be derived from or explained in terms of its underlying elements. This stipulation is necessary in order to prevent RP emergence from having the occult character of ontolog-

ical emergence. On the other hand, by condition (3), RP emergence differs from epistemological emergence in that an emergent real pattern is in a strong sense *autonomous* from its underlying elements. Recall that the Laplacean super-physicist Martians from Dennett's thought experiment have no need to appeal to epistemologically emergent phenomena in order to make their predictions. Everything that epistemologically emergent phenomena might allow them predict they can also predict from the laws of fundamental physics. Epistemologically emergent phenomena are nothing more than epistemic crutches that make computations more tractable for beings like us who have limited computational resources and power; the Martians have no need for them.

This, however, is not the case with RP emergence. Insofar as the Martians are blind to emergent real patterns, there are phenomena which they will be unable to predict, but which we who have access to the patterns *can* predict. Suppose, for example, that Bob is ill-tempered. This is a real pattern in his behavior, and we shall assume that it somehow emerges from the physiological and neural processes that take place in his body. On the basis of this pattern, I am able to predict, not only what he *will* do in a given circumstance, but also what he *would* do in an indefinite variety of possible circumstances. The Martians too will be able to predict what Bob will do in a given circumstance. Suppose that I decide to put some *wasabi* in his dessert. The Martians will be able to predict Bob's ensuing fit of anger, but they can do so only by tracking every physical condition that could conceivably have an effect on the outcome, such as the temperature, the direction of the wind, and perhaps whether a butterfly had been fluttering in a specific location in Brazil the day before. But they cannot predict what Bob *would* do in a merely hypothetical circumstance, because they are unable to isolate the factors relevant to the prediction (such as Bob's being ill-tempered) from those that are irrelevant. Since I know that Bob is ill-tempered, I can predict how he would act in an indefinite variety of possible circumstances, without having to specify conditions such as the temperature, the direction of the wind, etc. But the Martians cannot make their prediction without all of this information, which in fact is irrele-

vant to the prediction, and this is because (by hypothesis) they do not have access to the real pattern that Bob is ill-tempered. In order to know what factors are relevant to making a prediction, one must have access to the appropriate patterns.

Notice also that condition (3) gives RP emergence the kind of informational non-redundancy that Ladyman and Ross were aiming at in their definition of real patterns (§1.3). An emergent real pattern supports predictions that are impossible if we only have access to descriptions of its underlying elements, and in this sense it is *indispensable* without sacrificing projectibility. We are thus able to formulate the idea of non-redundancy without appealing to the problematic notion of "physically possible" computers.

Many philosophers seem to think that an emergent pattern embodies less information than descriptions of its underlying elements. In other words, they think that even if a "higher-level" description of a system is useful or perhaps indispensable in making predictions, it is merely the result of shaving away some of the information contained in the "lower-level" description. Even Dennett, whose theory of real patterns goes a long way towards offering a modern vindication of Scholastic realism, seems to lapse into this position when he characterizes abstract objects as "lossy compression[s]" (Dennett 2000:360). An emergent real pattern, however, can embody *more* information than descriptions of its underlying elements, in that it supports predictions which the latter do not.[29] And this is precisely because patterns are more *general* than the elements instantiating them. One is tempted to see in the view that emergent patterns are mere "compressions" of their underlying dynamics the "Ockhamistic prejudice" that Peirce refers to in the passage quoted in the epigraph of the present chapter: "Get rid, thoughtful Reader, of the Ockhamistic prejudice of political partizenship that in thought, in being, and in development the indefinite is due to a degeneration from a primal state of perfect definiteness" (CP 6.348, 1907).

[29] Interestingly, Erik Hoel (2017) has arrived at the same conclusion, that a macro-level description of a system can contain more information than a micro-level description, through an application of information theory to the analysis of causal structures; see Hoel (2018) for a non-technical exposition.

§3.4 Bedau on Weak Emergence

In this section I want to throw what I call RP emergence into sharper relief by comparing it with Mark Bedau's related notion of *weak emergence* (Bedau 1997, 2002). Bedau distinguishes three kinds of emergence: nominal, weak, and strong. A *nominally emergent* property is "a macro property that is the kind of property that cannot be a micro property" (Bedau 2002:9). This is simply an alternative formulation of the characterization of emergence given in §3.1. Nominal emergence is the barest and broadest notion of emergence, and encompasses both weak and strong emergence as special cases. Note that it corresponds to condition (1) in my definition of RP emergence. A *strongly emergent* property is one which, in addition to being nominally emergent, is a "supervenient propert[y] with irreducible causal powers" (Bedau 2002:10). These macro causal powers have a determinative influence on both the macro and micro levels, and in the latter case it is called *downward causation*. Although Bedau's notion of strong emergence is defined differently from what I have been calling ontological emergence, it is similar to the latter in that the causal powers associated with it are "brute" natural powers that arise inexplicably from the micro elements or processes (Bedau 2002:11).

Finally, a *weakly emergent* property of a system is one which, in addition to being nominally emergent, can be derived only through a step-by-step simulation of the system. Bedau's definition is as follows: "Assume that *P* is a nominally emergent property possessed by some locally reducible system *S*. Then *P* is weakly emergent if and only if *P* is derivable from all of *S*'s micro facts but only by simulation" (Bedau 2002:15). A *locally reducible* system is, roughly, a system whose macro properties are all structural properties—they are wholly constituted by the states and locations of the system's micro entities—and whose micro dynamics is context-sensitive in the sense that a micro entity's state depends on the states of its micro-level neighbors (Bedau 2002:14). Since weakly emergent properties are macro properties of a locally reducible system, they are wholly constituted by the states and locations of the system's micro entities. In other words, they are ontologically reducible to micro phenomena: "their existence

consists in nothing more than the coordinated existence of certain micro phenomena" (Bedau 2002:12). This is what makes weak emergence a weaker form of emergence than strong emergence, which involves irreducible causal powers. On the other hand, the context-sensitivity of the micro dynamics of locally reducible systems entails that understanding how the micro entities behave in isolation or in certain simple contexts does not, in general, enable us to predict how they will behave in more complicated contexts (Bedau 2002:14). Locally reducible systems thus possess a certain kind of unpredictability, and a weakly emergent property is a macro property of a locally reducible system that is unpredictable in a specific sense: it is underivable from knowledge about the micro entities except by explicit simulation of the system.

Although Bedau frames his definition of weak emergence in terms of *properties*, we can also define weak emergence in terms of *patterns*: Let *S* be a locally reducible system, and let *P* be a nominally emergent pattern of *S*, that is, *P* is a pattern that arises in *S* but cannot be manifested by the constituent elements of *S*. Then *P* is a weakly emergent pattern if and only if *P* is underivable from knowledge about the constituent elements of *S* except by explicit simulation of *S*. Hereinafter I will speak of weakly emergent patterns rather than properties.

What deserves emphasis is that the impossibility of deriving weakly emergent patterns except by explicit simulation is not a merely practical impossibility that might be overcome some time in the future, or by beings with greater computational power than humans. Weak emergence has nothing to do with the epistemic limitations of the human mind or lack of available computational resources. Rather, "it involves the formal limitations of any possible derivation performed by any possible device or entity" (Bedau 2002:17). To dramatize this point, Bedau considers a Laplacean supercalculator—not unlike the Martians we have been considering so far—whose computational speed and accuracy are not bounded by any human or hardware-related limitations. He insists that even such a being would not be able to derive weakly emergent patterns except by direct simulation (Bedau 2002:17). This is because the process leading up to the emergence of weakly emergent patterns is, to use the terminology of Wolfram (1985),

computationally irreducible, that is, as a matter of principle there can be no short-cut derivations of these patterns that are simpler than the natural computational process by which they are generated.

A good way to understand the notion of computational irreducibility is to consider a process that is *computationally reducible*. Suppose, for example, that we throw a stone straight up into the air. Let us assume that the stone is subject only to gravity and an air resistance proportional to its velocity. Given the initial position and velocity of the stone, we can determine, using Newton's laws of motion, the position and velocity of the stone at any desired time *t* after it has been thrown. Newton's laws thus provide us with a short-cut derivation of the system's state at time *t*: we do not have to actually go through the entire evolution of the system leading up to time *t* in order to determine the system's state at that particular time. This is what it means to say that a process is computationally reducible. A computationally *irreducible* process, on the other hand, cannot be bypassed in this way. It is of such complexity that in order to determine the state of the system at some time *t*, we must explicitly follow the entire evolution of the system leading up to *t*.

Using the notion of computational irreducibility, we can define weakly emergent patterns as follows: a pattern is said to be weakly emergent if the process leading up to its emergence is computationally irreducible. Note that this implies that weak emergence must be a form of diachronic emergence (§3.2). The emergence of patterns in the Game of Life, which we will discuss in detail below, is an example of a computationally irreducible process, and hence of weak emergence.

Even if the impossibility of deriving weakly emergent patterns except by explicit simulation is an impossibility in principle rather than an impossibility in practice, one might still urge that the impossibility is merely an epistemological one, and that weak emergence is therefore a form of what I have been calling epistemological emergence. The argument would go something like this: since, by definition, weakly emergent patterns are nothing more than aggregations of the

micro phenomena that constitute them, they do not have any real explanatory power: "all the explanatory power resides at the micro level and the macro phenomena are merely an effect of what happens at the micro level" (Bedau 2002:37). Hence, even if weakly emergent patterns have explanatory autonomy, this autonomy is a merely epistemological one—it amounts to nothing more than "our inability to follow through the details of the complicated micro causal pathways" (Bedau 2002:38)—and does not reflect any autonomous and irreducible feature of reality. Bedau's reply to this line of argument throws interesting light on the relation between weak emergence and RP emergence. In response to the argument, he makes a distinction between cases of weak emergence for which the argument is sound, and cases for which it is not. He grants that in some cases, weakly emergent patterns are indeed mere effects of what happens at the micro scale, and their explanatory autonomy is merely epistemological (Bedau 2002:38).

As an example, consider John Conway's Game of Life. This is a cellular automaton consisting of an infinite, two-dimensional lattice of square cells, each of which is in one of two possible states, dead or alive. Time in the Game of Life flows in discrete steps. At each time step, each cell updates its state according to a simple function of its own state and the states of its eight neighboring cells in the previous step. The update rule for the Game of Life is as follows:

> (1) A living cell stays alive if either two or three of its neighbors were alive in the previous step; otherwise it dies.
>
> (2) A dead cell becomes alive if exactly three of its neighbors were alive in the previous step; otherwise it remains dead.

Given a suitable initial configuration of living and dead cells, the above rule will generate various enduring patterns in the playing field; whether and what patterns appear depend on the initial configuration. For example, there is a particular pattern called a "glider" that moves diagonally across the field, shifting one cell along the diagonal every four time steps (Fig. 3.1). There

are also various kinds of "glider guns" that periodically shoot gliders; one particular type of glider gun, known as the Gosper glider gun, is shown in Fig. 3.2.

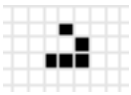

Fig. 3.1 Glider (the black cells represent living cells and the white cells represent dead cells)

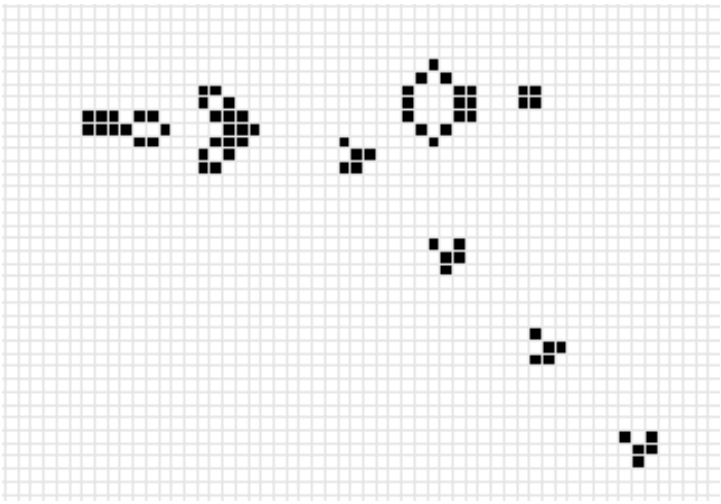

Fig. 3.2 Gosper glider gun

Even if there is no glider gun, gliders can also be produced accidentally from interactions among other patterns or patternless clusters of living cells. Bedau asks us to consider a "configuration in the Game of Life that accidentally … emits a stream of six evenly spaced gliders moving along the same trajectory" (Bedau 2002:38). He argues that the emergence of this glider stream is an example of merely epistemological weak emergence, because it has no overarching explanation: "The explanation for the glider stream is just the aggregation of the causal histories of the individual cells that participate in the process" (Bedau 2002:39). The glider stream is similar to the accidental succession of sixes in our example of throwing a die (§1.2): just as there is

no real law or *would-be* governing the succession of sixes, so there is no real law or *would-be* governing the glider stream.

On the other hand, suppose there is a glider gun shooting a stream of gliders, as in Fig. 2. As in the previous case, this glider stream can be explained by the aggregation of the causal histories of the individual cells participating in the process. However, there is more to this second glider stream: it is produced by a glider gun, which provides an overarching macro explanation of the stream. This macro explanation is applicable not only to the case at hand, but also to any other instance in which a glider gun shoots a stream of gliders. The aggregate micro explanation omits this information (Bedau 2002:39). Furthermore, the instances in which the macro explanation is applicable include counterfactual situations:

> The same glider stream would have been produced if the first six gliders had been destroyed somehow (e.g., by colliding with six other gliders). Indeed, the same glider stream would have been produced if the configuration had been changed into any number of ways, as long as the result was a gun that shot the same kind of gliders. Any such macro gun would have produced the same macro effect. (Bedau 2002:39)

The glider gun is thus autonomous from its underlying micro dynamics, because it is a macro pattern that "supports counterfactuals about what *would* happen in an indefinite variety of different micro situations" (Bedau 2002:41–42, emphasis mine). In other words, it is autonomous because it is an instance of RP emergence.

The accidental glider stream and glider gun are both instances of weak emergence, insofar as they are both underivable without actually going through the Game of Life step-by-step.[30]

[30] The underivability of patterns in the Game of Life without explicit simulation is a consequence of the fact that the Game of Life is Turing complete, that is, it can be used to simulate an arbitrary computer program. Suppose there is a general algorithm that allows us to accurately predict the behavior of the Game of Life for an arbitrary initial configuration, without going through it step-by-step. Since the Game of Life is Turing complete, this algorithm will also be able to determine the behavior of an arbitrary program with any possible input, including whether it will halt or not. But this contradicts the undecidability of the halting problem, so there can be no such algorithm.

Nonetheless, the former does not have the autonomy of the latter, and this is because the latter is projectible into counterfactual situations while the former is not. What this shows is that emergent patterns that are autonomous from their underlying elements are so by virtue of their being instances of RP emergence, not by virtue of being instances of weak emergence.

The accidental glider stream is an instance of weak emergence that is not an instance of RP emergence. Hence, not all instances of weak emergence are instances of RP emergence. Conversely, not all instances of RP emergence are instances of weak emergence. As an example of RP emergence that is not an instance of weak emergence, consider a thought experiment put forth by Hilary Putnam in "Philosophy and Our Mental Life" (Putnam 1975:295–97). Suppose we have a rigid board with two holes, a circle one inch in diameter and a square one inch high, and a cubical peg slightly smaller than one inch in each dimension. We want to explain the fact that the peg passes through the square hole but not the round hole. One way of going about would be to regard the board and peg as lattices of atoms, and attempt to calculate all the possible trajectories of the peg from the laws of elementary particle physics (if this sounds infeasible, we can suppose that the calculation is carried out by the Martians from Dennett's thought experiment). We could say that we have attained our explanation if we are able to deduce that the peg never passes through the round hole, but there is at least one trajectory in which it passes through the square hole.

There is, of course, a much simpler explanation. We simply note that both the board and peg are rigid, the round hole is smaller than the peg, and the square hole is larger than the cross section of the peg. This is an explanation that appeals to the *shape* of the holes and peg, which is a pattern that emerges from the way the atoms composing the board and peg are arranged. There is presumably nothing about this pattern that makes it underivable from knowledge about its underlying elements except by explicit simulation; it is a trivial result of the atoms being held together in a certain configuration. It is therefore not an instance of weak emergence. On the other hand, there is an indefinite variety of possible trajectories by which one could attempt to

make the peg pass through either of the holes, and the shape of the holes and peg allow us to predict, in the case of any of these possible trajectories, whether the peg would pass through (as well as explain why it passes through or not). Furthermore, the same kind of prediction (and explanation) will hold for any set of objects with the relevant geometrical features, regardless of their size, the material they are made of, etc. Just like the glider gun, the shape of the holes and peg supports predictions about what would happen in an indefinite variety of possible micro situations, including situations that are not actualized. It is thus an instance of RP emergence.

Let us recap: the computational irreducibility of weakly emergent patterns (i.e., the computational irreducibility of the process leading up to the emergence of weakly emergent patterns) gives these patterns a certain kind of unpredictability and explanatory autonomy. This makes weak emergence an "intermediate" type of emergence, stronger than merely nominal emergence but weaker than strong emergence. In this respect it is similar to RP emergence, which is also an "intermediate" type of emergence, stronger than epistemological emergence but weaker than ontological emergence. However, a weakly emergent pattern's computational irreducibility does not by itself guarantee that its explanatory autonomy is more than merely epistemological, reflecting an autonomous and irreducible feature of reality. In order for a pattern to have this kind of autonomy, it must be an instance of RP emergence, as illustrated by our example of Conway's Game of Life.

Chapter 4: Downward Causation and Teleology

> *When we speak of an "idea," or "notion," or "conception of the mind," we are most usually thinking—or trying to think—of an idea abstracted from all efficiency. But a court without a sheriff, or the means of creating one, would not be a court at all; and did it ever occur to you, my reader, that an idea without efficiency is something equally absurd and unthinkable? Imagine such an idea if you can! Have you done so? Well, where did you get this idea? If it was communicated to you viva voce from another person, it must have had efficiency enough to get the particles of air vibrating. If you read it in a newspaper, it had set a monstrous printing press in motion. If you thought it out yourself, it had caused something to happen in your brain.*
>
> Charles S. Peirce, *Minute Logic*

So far, building on Dennett's notion of real patterns and Peirce's modal analysis of generality, we have developed an account of RP emergence, showing how it is distinct from both epistemological and ontological emergence. I have further attempted to throw the concept of RP emergence into sharper relief by comparing it with Bedau's related notion of weak emergence. What we have not yet touched upon is the topic of downward causation. Downward causation—the idea that emergent phenomena exert a causal influence on their underlying elements—is often associated with ontological forms of emergence. I suggest that RP emergence could also involve this kind of causal power.

I will begin by motivating the idea of downward causation through a discussion of the phenomenon of *mutual entrainment* (§4.1). Then I will address two problems that seem to be inherent in the notion of downward causation, which I call the *incoherence problem* and *dispensability problem* (§4.2). Both have been outlined by Jaegwon Kim (1999, 2000).[31] I will then argue that these problems do not arise if we conceive of downward causation on the model of RP

[31] Kim (2000) is a reprint of the latter half of Kim (1999) with some minor changes and additions. When referring to a passage that appears in both versions, I will cite the page numbers of both versions.

emergence, in which case the downward cause—the "entity" that exerts the causal influence—is a general form rather than a concrete entity (§4.3). Next, I attempt to show that the causal influence associated with RP emergence is best understood as a form of *final causation* as conceived by Peirce (§4.4). I will also address several objections that may be levelled against the view I propose.

§4.1 Mutual Entrainment and the "Virtual Governor"

To begin with, let me motivate the idea of downward causation by discussing a phenomenon known as *mutual entrainment*. The term "entrainment" refers to the phenomenon whereby an oscillator synchronizes with an input signal. A familiar example is the circadian pacemaker or "biological clock," a biochemical oscillator that regulates the sleep-wake cycle of various organisms by "locking in" with the cycle of certain environmental cues, such as sunlight and temperature. Anyone who has experienced jetlag knows the disorienting effects of being thrown off of the natural sleep-wake cycle set by the circadian pacemaker.

Mutual entrainment is a synchronization that occurs spontaneously among mutually interacting oscillators. This phenomenon was first observed by the physicist Christiaan Huygens (1629–1695). In a 1665 letter to his father, he reported his observation that two identical pendulum clocks hung from a common beam synchronized with each other with a high degree of precision, but with the pendula swinging in opposite directions (Huygens 1893:233–34).[32] Initially he thought that this anti-phase synchronization was due to the stirring of the air caused by the motions of the pendula, but after a series of experiments he realized that it was due to the coupling of the clocks through the beam from which they hung.

Today we know that this "sympathy" between pendulum clocks is an instance of a more general phenomenon—mutual entrainment—that can be found in a wide range of settings throughout nature, from the synchronous flashing of fireflies in some parts of Southeast Asia, to

[32] See Pikovsky, Rosenblum, & Kurths (2001, Appendix A1) for an English translation of the letter.

the synchronous beating of the heart's pacemaker cells. While Huygens' experiment involved only two pendulum clocks, mutual entrainment can occur among a large number of mutually interacting oscillators. Thus, thousands of fireflies of certain tropical species are known to flash exactly in unison, and the millions of cells that constitute the sinoatrial node, the natural pacemaker of the human heart, produce rhythmic electric pulses in synchrony. What is interesting in both of these cases is that none of the individual elements has "knowledge" about the entire system. Rather, each element interacts only with its immediate neighbors, and yet an overall pattern emerges from these local interactions. It is almost as if there is something "governing" the behavior of the elements.

Norbert Wiener (1894–1964), the father of cybernetics, gave expression to this idea when he spoke of a system of mutually entrained oscillators being regulated by a "virtual governor" (Wiener 1965:201). This idea occurs in his discussion of an electric power grid, which is a network of AC generators. Each generator is an oscillator with a built-in regulator or governor that keeps its frequency within a comparatively narrow range. Although each generator does not produce a very steady output in isolation, when they are wired together, by virtue of their mutual feedback they produce a steady alternating current with an accuracy going far beyond that of any of the generators in isolation. Hence, it is as if the entire system is being regulated by single virtual governor. As E. M. Dewan points out, "[t]his virtual governor is not located in one spot in the system, but rather it pervades the system as a whole, so that it does not have a 'physical existence' in the usual sense. It is an *emergent property of the entire system* which goes far beyond what any single unit can accomplish in accuracy and power" (Dewan 1976:185). We can see the effect of this virtual governor by observing what happens when we add a new generator to a network of mutually entrained generators: the new generator will be pulled into synchrony with the oscillation frequency and phase of the overall system.

Dewan further suggests that perhaps the mind may be understood as something analogous to the virtual governor of this kind of power grid. As he puts it: "the 'virtual governors' of a power

grid stand in relation to the individual governors in a way which is analagous [*sic*] to the way consciousness and mind stand in relation to the activity of the neuronal units of the brain" (Dewan 1976:186). Thus, just as the virtual governor of a power grid emerges from the mutual interaction between individual generators and yet regulates the behavior of those governors, so the mind can be thought of as something that emerges from the mutual interaction between the neuronal units of the brain, and yet regulates the activity of those very units. Interesting as this analogy is, it should be pointed out that only those aspects of our mental life which play a role in governing our behavior—intentions, decisions, beliefs, habits, and so on—can be explained in this way; the purely "qualitative" aspects of consciousness—what if feels like to listen to a certain piece of music, for instance—as well as the brute fact that we experience the world at all, cannot be accounted for by the hypothesis that the mind is analogous to the virtual governor of a power grid.[33]

The regulating of a system of mutually entrained oscillators by a virtual governor is a prime example of the kind of causal influence that I have in mind when I speak of "downward causation" (though it should be noted that downward causation is by no means restricted to systems of mutually entrained oscillators). Now one might think that the real causal work is being done by the individual oscillators, and that it only *seems* there is a virtual governor regulating the system. In contrast, what I want to argue in the remainder of this chapter is that in the case of RP emergence, the emergent pattern is capable of exercising a real causal influence on its underlying elements, and that this influence consists in regulating the behavior of those elements, in the same way that a virtual governor of a system of mutually entrained oscillators regulates the behavior of those oscillators.

[33] In Peircean terms, we could say that only the mind's aspect of Thirdness (law, regularity) is an emergent phenomenon analogous to a virtual governor; its aspect of Firstness (pure feeling) and Secondness (brute actuality) are not emergent phenomena.

§4.2 Problems with Downward Causation

An early proponent of the idea that an emergent phenomenon can exert a causal influence on its underlying elements is the neurobiologist Roger W. Sperry (1913–1994). He gives the following example (Sperry 1969:534). Consider a situation where water molecules are being carried along by an eddy in a stream. The eddy is constituted by water molecules swirling around in a circular motion, and at the same time, the eddy causes the water molecules to move in just this way.

Biological systems provide striking examples of phenomena that seem to involve downward causation. Consider, for example, how the cells constituting an organ regulate their size and number in order to keep the entire organ at an "appropriate" size. Thus, if part of the liver is removed through surgery or injury, the remaining liver cells will increase their size (hypertrophy) and actively divide in order to replace the lost tissue. Interestingly, this regeneration process terminates when the liver has recovered its original size. It is as if the individual cells "know" when to stop increasing their size and dividing in order to maintain the size of the entire organ. Similarly, when a planarian is cut into pieces, each piece will regenerate, and this regeneration process ceases when each piece has become a complete organism. The biophysicist Kunihiko Kaneko has suggested that in such phenomena "the parts composing the whole are determined by the whole" (Kaneko 2006:27), and indeed, in examples like these it seems that there is some kind of downward causal influence regulating the behavior of the individual cells so as to maintain or bring about a certain global feature of the system.

Another putative example of downward causation often discussed in the literature is the influence of mental phenomena—intentions, beliefs, and the like—on the body. Suppose, for example, that I decide to go to the grocery store to buy some groceries. Let us further assume that this decision somehow emerges from the physiological and neural processes taking place in my body. This decision, despite having its basis in my body, will cause my body to behave in a certain way, namely, it will cause the muscles in my limbs and fingers to expand and contract in

such a way as to carry me to the front door of my house, lock the door, drive to the grocery store, and so on. A further putative example of downward causation is the prices of goods in a market economy. The price of a good emerges from interactions between buyers and sellers, but it also influences the actions of those very buyers and sellers.

The concept of downward causation, however, seems beset with difficulties. One potential problem is that the very notion of downward causation appears incoherent. There is certainly an air of paradox in maintaining that an emergent phenomenon is able to exert a causal influence on the very elements to which that phenomenon owes its presence. As Kim puts it: "Is it coherent to suppose that the presence of *X* is entirely responsible for the occurrence of *Y* (so *Y*'s very existence is dependent on *X*) and yet *Y* somehow manages to exercise a causal influence on *X*?" (Kim 1999:25, 2000:311). Let us call this the *incoherence problem* of downward causation.

Kim suggests that this problem can be circumvented by distinguishing between two types of downward causation: *synchronic downward causation* and *diachronic downward causation* (not to be confused with the distinction between synchronic emergence and diachronic emergence introduced in §3.2). Before formulating this distinction, however, a word on Kim's conception of emergence and downward causation is in order. For Kim, emergence pertains to parts and wholes, and accordingly, he conceives of downward causation as a causal influence which a whole exerts on one (or more) of its parts by virtue of its having a certain emergent property. Thus, he defines synchronic downward causation as the situation described below:

> At a certain time t, a whole, W, has emergent property M, where M emerges from the following configuration of conditions: W has a complete decomposition into parts $a_1, \ldots, a_n$; each a_i has property P_i; and relation R holds for the sequence $a_1, \ldots, a_n$. For some a_j, W's having M at t causes a_j to have P_j at t. (Kim 1999:28, 2000:314)

Diachronic downward causation, on the other hand, is defined as the situation described below:

> As before, W has emergent property M at t, and a_j has P_j at t. We now consider the causal effect of W's having M at t on a_j at a later time $t + \Delta t$. Suppose, then, that W's having M at t causes a_j to have Q at $t + \Delta t$. (Kim 1999:29, 2000:315)

The idea here is that synchronic downward causation is a causal influence which a whole, by virtue of its having a certain emergent property, exerts on one (or more) of its parts *at the same time* that those parts give rise to the emergent property in question; while diachronic downward causation is a causal influence which a whole, by virtue of its having a certain emergent property, exerts on one (or more) of its parts *at a later time* than the time at which those parts give rise to the emergent property in question.

Kim argues that the incoherence problem only applies to synchronic downward causation, and does not apply to the diachronic kind (Kim 1999:28–31, 2000:314–16). According to Kim, synchronic downward causation is incoherent because it violates what he calls the "causal-power actuality principle," a "metaphysical principle" which he formulates as follows:

> For an object, x, to exercise, at time t, the causal/determinative powers it has in virtue of having property P, x must *already* possess P at t. When x is caused to acquire P at t, it does not already possess P at t and is not capable of exercising the causal/determinative powers inherent in P. (Kim 1999:29, 2000:315)

Thus, if we assume that a whole W, by virtue of its having an emergent property M, causes one of its parts a_j to acquire a property P_j at time t, then it follows from the causal-power actuality principle that a_j does not already possess P_j at t. But if a_j does not already possess P_j at t, then it follows again from the causal-power actuality principle that a_j is not capable of exercising its determinative powers necessary in bringing about the emergence of M at t (Kim 1999:29, 2000:315). Of course, one could reject the causal-power actuality principle, but—Kim argues—insofar as one accepts it as a plausible principle of causation and determination, then one must reject synchronic downward causation.

While there is—according to Kim—no incoherence in the concept of diachronic downward causation, he levels a separate argument against downward causation as such.[34] I quote the entirety of his argument below:

> If an emergent, *M*, emerges from basal condition *P*, why can't *P* displace *M* as a cause of any putative effect of *M*? Why can't *P* do all the work in explaining why any alleged effect of *M* occurred? As you may recall, I earlier argued that any upward causation or same-level causation of effect *M** by cause *M* presupposes *M*'s causation of *M**'s lower-level base, *P** (it is supposed that *M** is a higher-level property with a lower-level base; *M** may or may not be an emergent property). But if this is a case of downward emergent causation, *M* is a higher-level property, and as such it must have an emergent base, *P*. Now we are faced with *P*'s threat to preempt *M*'s status as a cause of *P** (and hence of *M**). For if causation is understood as nomological (law-based) sufficiency, *P*, as *M*'s emergence base, is nomologically sufficient for it, and *M*, as *P**'s cause, is nomologically sufficient for *P**. Hence, *P* is nomologically sufficient for *P** and hence qualifies as its cause. The same conclusion follows if causation is understood in terms of counterfactuals—roughly, as a condition without which the effect would not have occurred. Moreover, it is not possible to view the situation as involving a causal chain from *P* to *P** with *M* as an intermediate causal link. The reason is that the emergence relation from *P* to *M* cannot properly be viewed as causal. This appears to make the emergent property *M* otiose and dispensable as a cause of *P**; it seems that we can explain the occurrence of *P** simply in terms of *P*, without invoking *M* at all. (Kim 1999:32, 2000:318–19)

The basic idea of this argument is that an emergent property will compete with its emergent base for causal influence over a lower-level event, and the emergent base, being more fundamental, will win the competition, making the emergent property otiose and dispensable as a cause of the lower-level event. Let us call the problem posed by this argument the *dispensability problem*.

---

[34] To be more accurate, Kim's argument against downward causation is restricted to the variety that he calls "reflexive downward causation" (Kim 1999:26, 2000:312). Kim makes a distinction between *reflexive* downward causation, in which "[s]ome activity or event involving a whole *W* is a cause of, or has a causal influence on, the events involving its *own* micro-constituents," and *nonreflexive* downward causation, "in which an event involving a whole causes events involving lower-level entities that are not among its constituents" (Kim 1999:26–27, 2000:312). However, I have been using the term "downward causation" to refer exclusively to what Kim calls "reflexive downward causation," because I am suspicious of the notion of "levels" involved in the definition of nonreflexive downward causation (see §3.2).

§4.3 Downward Causation and RP Emergence

As pointed out by Menno Hulswit (2005), much of the difficulty and confusion in discussions of downward causation stems from the fact that different authors have different metaphysical assumptions, often implicit, about causation in general and downward causation in particular. This is also true in the case of Kim. It seems to me that the problems he identifies in the notion of downward causation—the incoherence problem and dispensability problem—stem from the particular framework and language that he uses in formulating the idea of downward causation, namely, that of "parts," "wholes," and "emergent properties." Recall that for Kim, downward causation is a causal influence which a whole exerts on one (or more) of its parts by virtue of its having a certain emergent property. On this model, the downward cause—the entity that exerts the downward causal influence—is the "whole," which is a concrete entity.

I suggest that this model fails to adequately capture the idea of downward causation. A model better suited to understanding downward causation is that of RP emergence. As we saw in §1.4, a pattern is *general*: it may manifest itself in this or that concrete instance, but the pattern itself is not a concrete entity. It is a general *form*, and a real pattern in particular is a general form endowed with a power of making concrete things and events conform to it. In the case of an emergent real pattern, this power can be seen as a downward causal influence on the underlying elements.

Let us return to Kim's two problems, and see how they can be dealt with in the framework of RP emergence outlined above. Let us first consider the incoherence problem. The problem was that, in the case of synchronic downward causation, a whole, by virtue of its having a certain emergent property *M*, is supposed to cause one (or more) of its parts to acquire a property *P*, at the same time that this part (or parts) contributes to bringing about the emergence of *M* by virtue of having *P*. Framed in this way, synchronic downward causation certainly appears incoherent. Now let us rephrase the situation in the language of patterns: an emergent real pattern *P* causes its underlying elements to behave in a certain way at the same time that those elements constitute or give rise to *P*. There is nothing incoherent about this situation if we allow that an emergent real

pattern, such as the eddy in Sperry's example, need not be actualized in order for it to exert its causal influence. The causal power of the pattern can be thought of as consisting in its *tendency to bring about its own actualization* given the relevant conditions. That is, the pattern acts as an ideal end state towards which the constituent elements are compelled to tend (not unlike the ideal limit of inquiry in Peirce's convergence view of truth; see §2.3). As we have seen in the preceding chapters, a real pattern has the modality of a *would-be*, and as such, it has the power of manifesting itself whenever certain conditions are fulfilled. Thus, in a certain sense, the pattern is *present* even when it is not actually realized.

This is quite similar to what Claus Emmeche, Simo Køppe, & Frederik Stjernfelt (2000) call *weak downward causation*. Using the language of dynamical systems theory, they describe a weak downward cause as an *attractor* in phase space, i.e., "a set of points in phase space in which trajectories with many different initial conditions end" (Emmeche, Køppe, & Stjernfelt 2000:27). Attractors have the property of being stable under perturbations: even if a trajectory is thrown off of an attractor by an external force, it will return to the attractor as long as the perturbation is not too large. Thus, even if an organism contracts a disease, or is given a push, it will eventually return to its former state as long as the change is not too severe. Emmeche, Køppe, and Stjernfelt claim that the stability of an attractor is identical to the "governing" of the behavior of the system, which easily lends itself to being interpreted as a case of downward causation: "the physical perturbation is regulated by the biological attractor" (Emmeche, Køppe, & Stjernfelt 2000:28). Furthermore, in line with the view I have been expounding here, they argue that an "attractor is a general *type*, of which the single phase-space points in its basin will be *tokens*" (Emmeche, Køppe, & Stjernfelt 2000:29). I believe that Emmeche, Køppe, and Stjernfelt's notion of weak downward causation offers a compelling picture of many instances of downward causation. However, it should be noted that their account is restricted to those systems that can be modelled as dynamical systems. Their view also differs from mine in that they explicitly avoid speaking of weak downward causation as teleological (Emmeche, Køppe, & Stjernfelt 2000:29), whereas I prefer to inter-

pret the downward causal influence associated with emergent real patterns in light of Peirce's conception of final causation (to be discussed below in §4.4).

Returning to Kim's incoherence problem, Hulswit has argued that "Kim's rejection of synchronic downward causation is based on the presupposition that the only sort of causality is efficient causality," and that there is no incoherence in synchronic downward causation if it is regarded as formal, final, or material causation (Hulswit 2005:271). While I agree that Kim seems to assume that the only kind of causality is efficient causality, I believe that Hulswit's reasoning here is based on a misreading of Kim's argument. His reasoning is as follows (Hulswit 2005:271). The problem with synchronic downward causation, according to Kim, is that the idea of causation involves *transitivity*. But if causation is transitive, then synchronic downward causation would seem to entail a kind of self-causation, which is absurd. However, only efficient causality involves transitivity, and so there is no problem with synchronic downward causation if we regard it as formal, final, or material causation, which do not imply transitivity.

It is true that Kim mentions the problem of transitivity and self-causation. Citing Sperry's example of water being carried along by an eddy, Kim writes:

> The individual water molecules swirling in a circular motion together cause the eddy to occur, but, says Sperry, the eddy causes the water molecules to move around just this way. If causation is transitive, as it is standardly supposed to be, doesn't this mean that the motion of the water molecules causes itself? (Kim 2000:314)

However, as we saw above, Kim rejects synchronic downward causation *not* because it seems to entail self-causation, but because it violates the causal-power actuality principle. Indeed, there is nothing about the problem of self-causation that makes it peculiar to synchronic downward causation. If it were the ground on which Kim rejected synchronic downward causation, then he would have rejected diachronic downward causation as well (which he does not, at least at this point in his paper). In either case, the problem of self-causation does not arise because, properly speaking, the emergence relation is not a causal relation, and so, even if causa-

tion were transitive, there would be no causal chain from a part to the whole and back to the part again (Kim 1999:32, 2000:319). To repeat, Kim's argument against synchronic downward causation is not based on the problem of self-causation, and hence, Hulswit's claim that "Kim's rejection of synchronic downward causation is based on the presupposition that the only sort of causality is efficient causality" is accurate only to the extent that the said presupposition leads Kim to formulate downward causes as concrete entities rather than general forms or patterns. It is inaccurate if it is taken to mean that Kim's rejection of synchronic downward causation is based on the presupposition that the only sort of causality is one that involves transitivity.

Next let us turn to Kim's second problem, the dispensability problem. The problem, it may be recalled, is that the possibility of explaining a micro event in terms of micro causes makes an emergent phenomenon dispensable as a cause of the micro event. Against this it suffices to remark that an emergent real pattern does not cause this or that particular event. Rather, as a general form independent of its particular manifestations, it brings about a *general tendency* in the behavior of a system, and it is indispensable for the purpose of explaining this tendency.

Indeed, as pointed out by Hulswit, it was the observation that similar or identical patterns manifest themselves in nature in radically different settings—a phenomenon that physicists call *universality*—that led to the idea that "there must be some causal 'influence' which, contrary to efficient causal influence, is independent from the components of the system, and which explains the form the system takes" (Hulswit 2005:271). A striking example of universality often discussed in the literature is critical phenomena: systems as diverse as fluids and magnets exhibit the same scaling laws when they approach certain critical states. Another fascinating example is the universal constants (known as Feigenbaum constants, after the mathematical physicist Mitchell Feigenbaum) associated with the period-doubling bifurcations in certain chaotic systems.[35] Nature abounds in universality: consider the isochronism of the pendulum, i.e., the inde-

[35] See Strogatz (2015, §§10.6–7) for an elementary exposition of Feigenbaum's universality theory, including the renormalization technique that he used to obtain these constants.

pendence of the pendulum's period from such features as its mass, amplitude (as long as it is sufficiently small), and material composition; or how an isolated system will eventually reach thermal equilibrium, irrespective of what the component particles are and how they are moving. The virtual governor of a system of mutually entrained oscillators that we considered in §4.1 is also universal in this sense, since it is a pattern that manifests itself in a wide variety of settings, irrespective of the specific nature of the individual oscillators. This kind of *general* behavior is what the concept of downward causation is supposed to explain, not this or that particular event. As we will see below, this is a hallmark of the Peircean conception of final causation.

§4.4 Peircean Teleology

Teleological explanations are routinely employed not only in everyday life but also in biology. We say, for example, that "I went to the grocery store to buy groceries," or that "the function of the heart is to pump blood throughout the body." Yet, the ancient idea that there are natural processes which are not only seemingly but *genuinely* goal-directed—the idea, in other words, that there are final causes operative in nature—has fallen into disrepute in modern science and philosophy. The same was true in the early 20th century, when Peirce, going against the predominantly mechanistic outlook of his time, put forward his view of final causation operating in nature. His conception of final causation differs from the traditional Aristotelian conception in several respects, and is intimately bound up with his Scholastic realism. As I mentioned above, I believe that the downward causal influence exerted by emergent real patterns is best understood as a form of final causation in Peirce's sense. My aim in this section is therefore two-fold. First, I want to flesh out Peirce's view of final causation, highlighting how it ties together with the account of RP emergence and downward causation I have been developing so far; and second, I want to reply to several objections that might be levelled against my view.

Peirce offers the most sustained discussion of his view of final causation in the second chapter of his projected book, *Minute Logic*, printed in part in *The Essential Peirce* under the title

"On Science and Natural Classes" (EP 2:115–32, 1902). There, he characterizes final causation as follows:

> [W]e must understand by final causation that mode of bringing facts about according to which a general description of result is made to come about, quite irrespective of any compulsion for it to come about in this or that particular way; although the means may be adapted to the end. The general result may be brought about at one time in one way, and at another time in another way. Final causation does not determine in what particular way it is to be brought about, but only that the result shall have a certain general character. (EP 2:120, 1902)

This is contrasted with efficient causation, which Peirce characterizes as follows:

> Efficient causation, on the other hand, is a compulsion determined by the particular condition of things, and is a compulsion acting to make that situation begin to change in a perfectly determinate way; and what the general character of the result may be in no way concerns the efficient causation. (EP 2:120, 1902)

The two kinds of causation are by no means mutually exclusive. Rather, they work in tandem in any given causal process. Peirce illustrates this by the following example:

> For example, I shoot at an eagle on the wing; and since my purpose,—a special sort of final, or ideal, cause,—is to hit the bird, I do not shoot directly at it, but a little ahead of it, making allowance for the change of place by the time the bullet gets to that distance. So far, it is an affair of final causation. But after the bullet leaves the rifle, the affair is turned over to the stupid efficient causation, and should the eagle make a swoop in another direction, the bullet does not swerve in the least, efficient causation having no regard whatsoever for results, but simply obeying orders blindly. (EP 2:120, 1902)

A purpose or intention is a *general type*: when I intend to do something, I do not care in what particular way the intention is realized, as long as it is realized in one way or another. The same holds true for any other kind of final cause. Thus, a final cause can be said to be a general type that "governs" efficient causes: it influences efficient causes in such a way that their outcome will conform to the general type. It is easy to see how Peirce's view of final causation is closely allied with his Scholastic realism: that generals have the power of governing natural events and processes is a testament to their reality.

Peirce's notion of final causation also provides an excellent framework for understanding the downward causal influence involved in RP emergence. As was mentioned earlier, an emergent real pattern can be regarded a general form endowed with the power of making its underlying elements behave in conformity to it. The pattern acts as an ideal end state—an "attractor," to use the terminology of Emmeche, Køppe, and Stjernfelt—towards which the elements are compelled to tend. For the sake of illustration, consider again the example of my decision to go to the grocery store. My intention to go to the grocery store is a pattern in my behavior, a general form to which my individual actions are obliged to conform. After making the decision, my actions are carried out in conformity to this pattern, so as to realize the intended ideal end state: the muscles in my limbs and fingers expand and contract in such a way as to carry me to the front door of my house, lock the door, drive to the grocery store, and so on. This situation coincides perfectly with Peirce's description of final causation.

Let us now turn to some objections that might be levelled against the view presented here. One potential objection to the idea of final causes is that final causation is an influence exerted by a future event on the present, and therefore it is at odds with the currently standard view of causation, according to which a cause must temporally precede (or at least be simultaneous with) its effect. The idea that final causes are concrete future events seems to be widespread in contemporary philosophy. The philosopher Richard B. Braithwaite, for example, describes the "problem" of teleological explanation as follows:

> In a [normal] causal explanation the explicandum is explained in terms of a cause which either precedes or is simultaneous with it: in a teleological explanation the explicandum is explained as being causally related either to a particular goal in the future or to a biological end which is as much future as present or past. It is the reference in teleological explanations to states of affairs in the future, and often in the comparatively distant future, which has been a philosophical problem ever since Aristotle introduced the notion of 'final cause' … (Braithwaite 1953:324)

As we have seen, however, on the Peircean view, final causes are not concrete future events; they are general types in the present. In other words, final causes are not future actualities but

present generalities. Therefore, there is no "backward" causal influence involved in final causation, and the "problem" identified in the above passage simply does not arise.

Another potential objection to my view is that positing final causes capable of influencing natural events and processes entails a violation of physical laws. Such an objection has been voiced by Kim in the context of downward causation. Citing an excerpt from Sperry, where he discusses how the vital and mental properties of an organism—aims, wants, needs—can influence the motion of the molecules composing the organism, Kim writes:

> This is an instance of what has been called "downward causation." The idea is that when certain wants and needs, aided by perceptions, propel a bird through the air, the cells and molecules making up the bird's body, too, are propelled, willy-nilly, through the air by the same wants, needs, and perceptions. If you add to this the further thesis, as Sperry would, to the effect that these psychological states and processes, though they "emerge" out of biological and physicochemical processes, are distinct from them, you are apparently committed to the consequence that *these "higher-level" mental events and processes cause lower-level physical laws to be violated*, that the molecules that are part of your body behave, at least sometimes, in ways different from the way they would if they weren't part of a living body animated by mental processes. (Kim 1992:120, emphasis in original)

First of all, what needs to be pointed out here is that the italicized sentence and its explanatory rewording are completely different claims. For clarity, let us number these claims:

> (1) "Higher-level" mental events and processes cause lower-level physical laws to be violated.
>
> (2) The molecules that are part of an organism's body behave, at least sometimes, in ways different from the way they would if they weren't part of a living body animated by mental processes.

Claim (2) in no way implies claim (1). The situation described in (2) happens all the time. For example, viruses behave differently when they are outside of their host cell and when they are in contact with their host cell (and hence "part of an organism's body"), but this evidently does not

entail any violation of physical laws. What (2) states is simply a particular case of the truism that the behavior of a physical system is in general sensitive to the context or environment in which the system is placed. Second, the scenario described by Sperry—where the aims, wants, and needs of an organism exert a downward causal influence on the motion of the molecules composing the organism—entails (2) but not (1). When a cell or molecule is part of an organism's body, it is subject to the regularities governing the organism, *in addition to* the physical laws that already govern it; again, this entails no violation of physical laws. Therefore, Kim's inference from Sperry's example to the italicized consequence (1) is totally unwarranted.

Against this one might further urge that if we assume the physical laws to be deterministic, then emergent real patterns such as the regularities governing an organism will have no causal role to play, since every event would be uniquely determined by the physical laws. The emergent patterns would be mere epiphenomena, lacking in any real causal efficacy. Now I do not believe that the actual physical world is deterministic. I agree with Peirce that there is an element of genuine chance in the universe.[36] However, since real patterns emerge not only in the actual physical world but also in certain computational and mathematical systems that obey simple, deterministic rules—such as cellular automata and certain dynamical systems—I will try to answer the objection on the assumption of determinism.

Recall that according to Peirce, efficient causation and final causation do not preclude each other. Rather, they can be seen as two different aspects of the same causal process. My reply will be based on this insight. Consider once again the example of my decision to go to the grocery store, and let us assume, for the sake of argument, that every movement of my body is precisely determined by deterministic physical laws. This situation can be described from two different standpoints. On the one hand, it can be viewed as a purely mechanical process, determined by the action of physical laws. From this standpoint, my decision to go to the grocery

[36] See Peirce's 1891 paper, "The Doctrine of Necessity Examined" (EP 1:298–311, W 8:111–25), the second installment of the *Monist* Metaphysical Series.

store and the subsequent movements of my body are wholly an affair of efficient causation. This would be a perfectly valid description of the process, but this does not preclude the possibility of viewing the same situation as an *intentional* process. From this standpoint, my decision to go to the grocery store and the subsequent movements of my body are an affair of final causation. Both of these accounts are perfectly legitimate ways of describing the same process. More importantly, we would have failed to notice something perfectly real if we did not see the process as an affair of final causation (just like the Martians in Dennett's thought experiment, see §3.1). Thus, generally speaking, emergent real patterns will have a genuine causal role to play even in situations where the pattern's constituent elements are subject to deterministic laws.

Lastly, one might object to the attribution of final causation to natural processes on the grounds that it is anthropomorphic. The idea that there are goal-directed processes in nature, it might be argued, involves an objectionable transfer of human qualities, such as intent, purpose, deliberation, or consciousness to non-human processes.[37] First of all, what needs to be emphasized in this connection is that on the Peircean view, a final cause is not necessarily a *purpose*: "A purpose is merely that form of final cause which is most familiar to our experience" (EP 2:120, 1902). There is no reason to believe that natural processes under the influence of a final cause somehow have a conscious goal or purpose as humans often do. With that being said, let me give two replies to the above objection. The first is due to Peirce, and the second is my own.

Peirce often insists that there is nothing wrong with an idea's being anthropomorphic: the charge that a certain conception is "unscientific because anthropomorphic" is "an objection of a very shallow kind, that arises from prejudices based upon much too narrow considerations" (EP 2:152, 1903). Anthropomorphism is simply unavoidable, not only because all of our conceptions ultimately derive from human experience, but more importantly, because the very possibility of scientific explanation is rooted in the hypothesis that there is some sort of analogy be-

[37] The biologist Ernst Mayr, in a classic paper on teleological explanation, lists this as one of the four objections that an acceptable teleological language must be immune to (Mayr 1974:94).

tween human reasoning on the one hand and the way the world is on the other: "every scientific explanation of a natural phenomenon is a hypothesis that there is something in nature to which the human reason is analogous; and that it really is so, all the successes of science in its applications to human convenience are witnesses" (EP 2:193, 1903); "there is in the being of things something which corresponds to the process of reasoning" (RLT 161, 1898). Therefore, anthropomorphism, far from being problematic, simply cannot be avoided.

Another reply to the charge of anthropomorphism is the following. It is evident that we humans are under the influence of final causes, whenever we intend to do something or act in order to fulfill a purpose. To deny that would be to deny the patently obvious. As I have argued above, even if it were possible to describe our intentional or purposeful behavior in a purely bottom-up fashion—say, in terms of neural and physiological processes—in no way would it follow that our intentions and purposes are mere epiphenomena, that we are not really governed by final causes. The only thing this view—that intentions and purposes are mere epiphenomena—has to recommend itself is its ontological parsimony, and this is not enough to compensate for the violence it does to our everyday beliefs and modes of thinking. Note that I am not claiming that our everyday intuitions should not be overridden by evidence and reason; far from it. However, the rejection of an ingrained intuition should be made on the basis of evidence and reason, and one's mere preference for desert landscapes does not qualify as such.

Let us grant, then, that humans are governed by final causes. Now humans are natural phenomena; we are a part of nature. Why should we suppose that humans (and perhaps some animals that are similar to humans) are the only natural phenomena governed by final causes? The idea that humans (and animals similar to humans) are *special* natural phenomena, in the sense that only *we* are capable of genuinely goal-directed behavior—in short, the idea involved in the charge that ascribing human qualities to non-human processes is objectionable—is, I contend, an instance of *anthropocentrism*. If we thoroughly and consistently pursue the idea that humans

are natural phenomena, and that we do not occupy any special place in the natural world, then we will be compelled to accept that final causation is a ubiquitous feature of nature.

Concluding Remarks

Let us recapitulate what we have achieved so far. In Chapter 1 we studied the basic properties of patterns, taking Dennett's theory of real patterns as our point of departure. We introduced Dennett's distinction between real and non-real patterns in terms of their predictive power, and further examined Ladyman and Ross's non-redundancy criterion. It was shown that a non-redundancy criterion is indeed necessary for defining the reality of patterns, but Ladyman and Ross's definition, which appeals to the notion of "physically possible computers," is problematic. We saw later in Chapter 3 that the non-redundancy of real patterns can instead be adequately captured by the concept of RP emergence. Finally, we concluded Chapter 1 by showing that patterns are *general* in the sense that they are multiply instantiable.

In Chapter 2 we delved into Peirce's theory of generality as encapsulated in his pragmatism and Scholastic realism. After laying down the terminology and framework of the problem of universals, we traced the development of Peirce's ideas on generality in a roughly chronological order, starting with his "On a New List of Categories," then his 1871 Berkeley review, and finally his papers on the pragmatic maxim. We introduced Peirce's mature view that a general has the modality of a *would-be*, in the sense that to ascribe a general to some particular object x is to recognize that x is governed by a series of laws or regularities that dictate not only how x *will* behave, but also how it *would* behave in certain kinds of counterfactual situations. Finally, we took up one of Peirce's late arguments for Scholastic realism, articulated in his 1903 Harvard Lectures on Pragmatism, and saw how the basic idea of this argument is strikingly similar to Dennett's definition of the reality of patterns in terms of their predictive power.

In Chapter 3 we explored the concept of emergence. After outlining the basic properties of what are commonly referred to as emergent phenomena, we introduced the distinction, often made in the literature, between epistemological and ontological emergence, and defined the views that I call ontological reductionism and radical emergentism. It was shown that both of

these positions are problematic and that both can be avoided if there is a third form of emergence that is neither epistemological nor ontological. After a preliminary discussion of the notion of levels, we introduced the concept of RP emergence, drawing on the ideas developed in the previous two chapters. We saw that the pattern that emerges in any instance of RP emergence is in a strong sense autonomous from its underlying elements, making this type of emergence distinct from epistemological emergence, without on the other hand collapsing it into ontological emergence. Finally, I attempted to throw the concept of RP emergence into sharper relief by comparing it with Bedau's related notion of weak emergence.

In Chapter 4 we discussed the concept of downward causation, often associated with ontological forms of emergence. We began by motivating the idea of downward causation through a discussion of the phenomenon of mutual entrainment. Next we addressed two problems that seem to be inherent in the concept of downward causation—the incoherence problem and dispensability problem—both of which have been outlined by Kim. It was then shown that these problems do not arise if the downward cause—the entity that exerts a downward causal influence—is regarded as an emergent real pattern rather than a concrete entity as in the case of Kim. Finally, we saw that the downward causal influence associated with emergent real patterns is best understood as a form of final causation in Peirce's sense.

My aim in this dissertation has been twofold. The first is to elucidate the issues surrounding the concept of emergence by approaching the topic from the standpoint of *patterns*, and employing the resources of Peirce's philosophy to do so. The basic idea of my approach has been that patterns are generals, and hence, that the question of their reality is a variation on the traditional problem of universals. I hope I have been able to show that by applying this idea, we can bring some clarity to the issues surrounding the concept of emergence, such as whether all instances of emergence are epistemological, and whether there is such a thing as downward causation—issues which, in my opinion, have often been obfuscated by nominalist assumptions and tendencies.

My second aim has been to update Peirce's Scholastic realism in the form of a realism about patterns. While it is true that much more work needs to be done in weaving together and making sense of the grand system of philosophy that Peirce envisioned, philosophers should above all else be wary of overspecializing and losing contact with the living sciences. As I noted in the introduction, we are seeing today rapid advances in our knowledge of complex systems and self-organizing phenomena, as well as the advent of new techniques in machine learning. To this we can add the development of algorithmic information theory, which, as we saw in §1.1, is intimately bound up with the concept of patternhood. It is only fitting, then, that Peirce's philosophy be updated in light of these developments, just as Peirce himself transmogrified the traditional problem of universals in light of the scientific developments of his day. In so doing we are providing a demonstration of the abiding relevance of Peirce's philosophical ideas in our efforts to make sense of some of the foundational issues in today's science.

Supplementary Chapter: Peirce's "New List of Categories"

In this supplementary chapter I provide a section-by-section commentary of Peirce's "On a New List of Categories" (EP 1:1–10, W 2:49–59; hereinafter referred to simply as the "New List"), which was discussed in §2.2 of this dissertation. Peirce's "New List," presented in 1867 to the American Academy of Arts and Sciences and published in its *Proceedings* the following year, is the first publication in which he identifies and derives his three universal categories, which in his later works he will call Firstness, Secondness, and Thirdness. The importance of this work has been widely recognized by commentators. Murray Murphey writes that among all of Peirce's published papers there is none "so important in its content" (Murphey 1993:66), while Donald Buzzelli holds that the paper "is a foundational one for Peirce's entire philosophy" (Buzzelli 1972:63). Indeed, the "New List" is not only the culmination of Peirce's early efforts to identify and derive the universal categories; it is also the first publication in which he gives a definition of *interpretant*, introduces the threefold division of signs into icons (likenesses), indices, and symbols, puts forward the idea of predication as hypothetical, offers a sustained discussion of the mode of mental separation that he calls *prescision* (which will also play an important role in Peirce's late phaneroscopic derivation of the categories), and presents his reformulation of the Roman/medieval trivium. These are all recurring themes in Peirce's later writings.[38]

Despite its importance, the "New List" is also notorious for its difficulty. Murphey says of it: "Certainly of all Peirce's published papers there is none which is so cryptic in its statement of

[38] It is not my intention here to weigh in on the debate of whether the "New List" is essential to understanding Peirce's philosophy as a whole. T. L. Short (2013) has argued, persuasively in my opinion, that this is not the case, and that the claim that the "New List" is the "keystone of Peirce's system of philosophy" (made by the editors of *The Essential Peirce*; see EP 1:1) is misleading. Nonetheless, I believe it will be agreed by all parties that the "New List" is crucial in understanding how Peirce's philosophy developed throughout his career, and that it is a worthwhile attempt to try to make sense of the argument presented in the paper.

essentials, so ambiguous in its definition of terms, so obscure in its formulation of the central doctrine …" (Murphey 1993:66). In this chapter I will attempt to bring some clarity to this paper by following Peirce's argument in the form of a section-by-section commentary. I will try to clarify points of obscurity and will not hesitate to advance and defend my own reading where there are issues of interpretation. Of particular note are my interpretation of Peirce's method of deriving the categories, and my interpretation of what he means by "reference to a correlate," which constitutes the second (intermediate) category.

The "New List" consists of fifteen sections. My exposition will be restricted to §§1–13, which is where Peirce presents his derivation of the categories (the first half of §14 will also be taken up in my discussion of the *correlate*). As noted in the introduction, Peirce's theory of the categories first set out in the "New List" constitutes the undercurrent of many of the ideas developed in this dissertation. In particular, we shall see how the theory of cognition that Peirce presents in this paper implies that regularity is the basis of cognizability, as suggested in the introduction of this dissertation.

§S.1 Opening sections (§§1–2)

§1 of the "New List" is as follows:

> §1. This paper is based upon the theory already established, that the function of conceptions is to reduce the manifold of sensuous impressions to unity, and that the validity of a conception consists in the impossibility of reducing the content of consciousness to unity without the introduction of it. (EP 1:1, W2:49)

This opening section should be understood as specifying the theoretical framework within which Peirce will be working in the "New List." The language used is unmistakably Kantian: "manifold" is the English rendering of Kant's *das Mannigfaltige*, which refers to the multiplicity of sense impressions before it has been ordered into a unified cognition by the application to it of conceptions. It is, roughly speaking, the "raw material" or "stuff" of experience which must

be brought into a unified form by the synthesizing function of conceptions in order to produce a cognition—this process Peirce calls the "reduction" of the manifold, or "reduction" of the content of consciousness, to unity. The above passage should be further understood as providing a functional definition of the key term *conception*, as well as a definition of what it means for a conception to be *valid*: a conception is precisely that whose function is to reduce the manifold of sensuous impressions to unity, and it is said to be *valid* if it is indispensable in enacting this reduction.

What is the "theory already established" that Peirce is referring to? The editors of *The Essential Peirce* have added a note citing Book 1 of the Transcendental Analytic in Kant's *Critique of Pure Reason* (EP 1:373). This book consists of two chapters: the first is the "Guide" to the categories (often referred to as the "metaphysical deduction" of the categories), and the second is the transcendental deduction of the categories. That the "theory already established" refers to the Transcendental Analytic has been disputed by T. L. Short (2013:277–83), on the grounds that Peirce replaces Kant's metaphysical deduction with a different one in the "New List," and that there is strong evidence that he rejects the need for a transcendental deduction. Short suggests instead that the "theory already established" refers to Lecture VIII of Peirce's own 1866 Lowell Lectures (Short 2013:279).

As Short notes, there is no surviving text of Lecture VIII, so we do not know what Peirce argued in that lecture. Furthermore, Short himself admits that the supposition that the "theory already established" refers to Peirce's eighth Lowell Lecture is "implausible," but says that he "can think of no hypothesis more plausible" (Short 2013: 279). However, it seems to me more plausible to assume that the "theory already established" does indeed refer to Kant, as indicated by the editors of *The Essential Peirce*. A point that deserves notice is that Peirce explicitly cites Kant in his 1894 rewriting of the "New List" (R 403), intended as the opening chapter of his complete but unpublished book *How to Reason: A Critick of Arguments*. In the 1894 version of §1 Peirce writes: "Kant, the father of modern philosophy, said that the function of conceptions

is to reduce the manifold of sensuous impressions to unity" (R 403:2, 1894). This, I believe, is sufficient evidence that the "theory already established" of the 1867 version is a veiled reference to the German master.

While Peirce certainly does not share every view set forth by Kant in the *Critique* or the Transcendental Analytic in particular, his attempt to demonstrate the validity of the categories by showing that they are *indispensable* in reducing the manifold of sensuous impressions to unity is precisely what Kant sets out to do in the transcendental deduction—note that Kant uses the term "deduction" not in the sense of a necessary inference, but in the legal sense of a proof establishing a claim of legitimacy, in this case the legitimacy of the use of certain concepts. Indeed, as we will see below, the method of prescision that Peirce uses to justify his derivation of the categories can be regarded as a *transcendental* method à la Kant, in that it shows whether a given conception is a condition for the possibility of introducing another conception, and ultimately of experience in general.[39] Hence it is only natural that Peirce should cite Kant in specifying the theoretical framework within which he will be working.

Let us now turn to §2. The entirety of this section is as follows:

> §2. This theory gives rise to a conception of gradation among those conceptions which are universal. For one such conception may unite the manifold of sense and yet another may be required to unite the conception and the manifold to which it is applied; and so on. (EP 1:1, W2:49)

By "those conceptions which are universal" Peirce means the categories.[40] He explains the term "universal" in Lecture IX of the 1866 Lowell Lectures as follows: "Of the numerous concep-

[39] In interpreting Peirce's prescision as a transcendental method, I am following Gava (2011). See also Kemling (2018) for an overview of the debate on Peirce's method of deriving the categories in the "New List."

[40] It should be noted, however, that Peirce will later abandon the idea that the categories are conceptions, and instead suggest that they are rather "moods or tones of thought" (EP 1:247, 1887–88). Later still, he will come to describe them as "not concepts but merely *elements* of concepts—what

tions of the mind, some apply only to certain special collections of impressions and are called *particular*. Others apply to all collections of impressions and are called *universal*" (W 1:473, 1866). Thus, the conception *table* is particular and not universal because it only applies to certain special collections of impressions, namely those produced by a table. A universal conception or category, on the other hand, is one that applies to all collections of impressions—in other words it is a conception that is operative in every act of cognition whatsoever. Peirce's aim in the "New List" is to identify these universal conceptions and thereby explicate the logical structure of the cognitive process at its most fundamental level.

Noteworthy here is the idea that the categories form a "gradation"; this is a significant departure from Kant. As the result of his derivation Peirce will obtain five categories, arranged in a hierarchical order of increasing abstractness. In the order of most abstract to least abstract ("farthest from sense" to "nearest to sense"), the categories are as follows (EP 1:6, W 2:54):

Being
    Quality (Reference to a Ground)
    Relation (Reference to a Correlate)
    Representation (Reference to an Interpretant)
Substance

In later writings *being* and *substance* will be dropped from the list of categories, and the three intermediate categories *quality*, *relation*, and *representation* will come to be called Firstness, Secondness, and Thirdness, respectively.

§S.2 Substance (§3)

Next let us turn to §3. Here Peirce introduces the category "nearest to sense," the conception of *substance*. The section is reproduced in its entirety below:

---

fluorine was among chemical substances until Moissan isolated it. Or better like *ions*" (RL 387b:328, 1908).

> §3. That universal conception which is nearest to sense is that of the *present, in general*. This is a conception, because it is universal. But as the act of *attention* has no connotation at all, but is the pure denotative power of the mind, that is to say, the power which directs the mind to an object, in contradistinction to the power of thinking any predicate of that object,—so the conception of *what is present in general*, which is nothing but the general recognition of what is contained in attention, has no connotation, and therefore no proper unity. This conception of the present in general, of IT in general, is rendered in philosophical language by the word "substance" in one of its meanings. Before any comparison or discrimination can be made between what is present, what is present must have been recognized as such, as *it*, and subsequently the metaphysical parts which are recognized by abstraction are attributed to this *it*, but the *it* cannot itself be made a predicate. This *it* is thus neither predicated of a subject, nor in a subject, and accordingly is identical with the conception of substance. (EP 1:1–2, W 2:49)

Peirce's derivation of the categories in the "New List" proceeds through an analysis of the cognitive process, which in turn is modelled as a process of *predication*. This process is set in motion by an encounter with something (the *present*) which calls for explanation. This *something* requires explanation because at this initial stage, where it still has not undergone predication, we cannot say anything about *what* it is or *how* it is. We can only direct our attention to it, and recognize it as an *it*. This *it* or the *present* is what Peirce calls the *substance*, drawing on the definition of the term given by Aristotle in the *Categories*: Peirce's statement that the *it* is "neither predicated of a subject, nor in a subject" is a deliberate echo of Aristotle's definition of *substance* (*οὐσία*) in the *Categories* V, 2a13, as "that which is neither predicated of a subject, nor in a subject." Note, however, that the substance itself is not a category. The first category is the *conception of* substance, or the *conception of* the present, in general: notice that Peirce writes that the "universal conception which is nearest to sense is *that of* the *present, in general*" (emphasis added to "that of"). The substance itself is only a bare *it* and is not a conception. The conception of substance is the conception at work whenever we recognize the present as something present (and something in need of explanation)—it is, as Peirce writes, "the general recognition of what is contained in attention"—and is the first universal conception that sets the cognitive process in motion.

§S.3 Being (§4)

Next, in §4, Peirce takes up the conception of *being*, which is the last universal conception that completes the cognitive process. Here I will reproduce only the first part of this section:

> §4. The unity to which the understanding reduces impressions is the unity of a proposition. This unity consists in the connection of the predicate with the subject; and, therefore, that which is implied in the copula, or the conception of *being*, is that which completes the work of conceptions of reducing the manifold to unity. (EP 1:2, W2:49–50)

According to Peirce, the cognitive process comes to an end with the formation of a proposition. To put it metaphorically, the cognitive process starts with a question mark, the substance, and ends with a period, the proposition. A proposition is formed by connecting a predicate to a subject. That which initially manifested itself as something in need of explanation, which corresponds to the subject, is explained, and its initial confusedness removed, by the application to it of a predicate. The reduction of the manifold of sensuous impressions to the unity of a proposition is thus achieved. Peirce says that since a proposition consists in the connection of the predicate with the subject, the conception implied in the copula, which he calls *being*, is the last universal conception which brings the cognitive process to an end (or more accurately, a temporary halt).

A question that arises at this point is whether Peirce is here considering only predications involving one logical subject, or also predications involving multiple logical subjects. In other words, one may wonder whether Peirce's account of predication in the "New List" is restricted to monadic predicates, or also embraces relational predicates. My view is that it embraces relational as well as monadic predicates. It is true that throughout the "New List" Peirce uses the word "subject" in the singular. However, in §15 he speaks of "[t]he objects indicated by the subject (which are always potentially a plurality,—at least, of phases or appearances)" (EP 1:9, W 2:57–58), and goes on to discuss the following argument (EP 1:9, W 2:58):

Whatever is the half of anything is less than that of which it is the half;
A is half of B:
∴ A is less than B.

He further remarks that "[t]he subject of such a proposition is separated into two terms, a 'subject nominative' and an 'object accusative'" (EP 1:9, W 2:58).

It is remarkable that already in 1867, when Peirce has still not worked out his logic of relatives or relations, he is dealing with arguments involving relational predicates—arguments that cannot be handled within the limitations of the traditional Aristotelian logic. Thus, contrary to a claim often made by commentators (e.g. Murphey 1993:152–53; Short 2013), Peirce's analysis of propositions in the "New List" is not restricted to those of monadic subject-predicate form. As we will see below, in the case of a relational predication, the correlate will play the role of the second object (that with which the initial substance is in relation).

On the other hand, it should be noted that at this stage he does not seem to have in mind relations involving more than two subjects. This is borne out by a statement that Peirce later makes in 1898, referring to his 1867 list of categories:

> I now undertook to ascertain what the conceptions were. This search resulted in what I call my categories. I then named them Quality, Relation, and Representation. But I was not then aware that undecomposable relations may necessarily require more subjects than two; for this reason *Reaction* is a better term [than *Relation*]. (CP 4.3, 1898; emphasis in original)

It seems that it is not until his study of the logic of relatives (c.1870) that Peirce comes to realize the irreducibility of triadic relations to monadic and dyadic ones; see his 1870 paper "Description of a Notation for the Logic of Relatives," where he argues that there are "three grand classes" of relative terms (W 2:364–65).

§S.4 The three modes of mental separation (§5)

The argument so far has established that substance and being are respectively the beginning and end of all cognitive processes. The next step is to search for any intermediate categories that may lie between these two, and thereby "retrace the path that the consciousness travels each time it pronounces a synthetic judgment" (De Tienne 1996:193).[41] The following section, §5, is where Peirce lays out his method of *prescision*, a mode of mental separation that will play a crucial role in his derivation of these intermediate categories. Here he distinguishes three modes of mental separation: *discrimination*, *prescision*, and *dissociation*. These are defined as follows (EP 1:2–3, W 2:50–51):

> **Prescision** (**abstraction**): The mental separation "which arises from *attention to* one element and *neglect of* the other. Exclusive attention consists in a definite conception or *supposition* of one part of an object, without any supposition of the other."
>
> **Discrimination**: "Discrimination has to do merely with the senses of terms, and only draws a distinction in meaning."
>
> **Dissociation**: "Dissociation is that separation which, in the absence of a constant association, is permitted by the law of association of images. It is the consciousness of one thing, without the necessary simultaneous consciousness of the other."

According to Peirce, "[a]bstraction or prescision … supposes a greater separation than discrimination, but a less separation than dissociation" (EP 1:3, 2:50)." This statement should be understood to mean that the conditions under which prescision is possible are stronger than those under which discrimination is possible, but weaker than those under which dissociation is possible. In other words, prescision is more demanding than discrimination, but less demanding than dissociation. The relative "strength" of the three modes of separation is illustrated by the following example (see also Fig. 1):

> Thus I can discriminate red from blue, space from color, and color from space, but not red from color. I can prescind red from blue, and space from color (as is manifest from the fact that I actually believe there is an uncolored space between my face and the

[41] De Tienne's French text will be cited in English translation; all translations are my own.

wall); but I cannot prescind color from space, nor red from color. I can dissociate red from blue, but not space from color, color from space, nor red from color. (EP 1:2–3, W 2:50–51)

	blue without red	space without color	color without space	red without color
Discrimination	◯	◯	◯	×
Prescision	◯	◯	×	×
Dissociation	◯	×	×	×

Fig. 1 Table illustrating the relative "strength" of the three modes of mental separation, adapted from Peirce's preliminary draft of the "New List" (W 1:519, 1866)

Discrimination, Peirce writes, "has to do merely with the senses of terms, and only draws a distinction in meaning." A more detailed explanation of this mode of separation is given in Peirce's 1894 rewriting of the "New List." In the rewritten version of §5 he explains discrimination as follows:

> *Discrimination* is a mere distinction of meaning. Thus, it is impossible to suppose there is color, without supposing there is a surface. Accordingly, although we can readily suppose the *sensation* of color to exist without any idea of space, yet color, as something objective, in the sense in which we understand it, cannot be supposed without three dimensions, at least. But we can perfectly well *discriminate* color from space; for this merely consists in recognizing that color involves something not necessarily involved in the supposition of space. (R 403:5, 1894)

The idea here seems to be this: two conceptions can be discriminated iff one of them contains in its meaning something not contained in the meaning of the other, i.e., if they are semantically distinct. Thus, color can be discriminated from space (we can have an idea of color without space) because color involves something not contained in the meaning of space, whereas red cannot be discriminated from color (we cannot have an idea of red without color) because the conception of color does not contain anything over and above that of red, nor does the conception of red contain anything over and above that of color—being a color is all there is to being

red.[42] The difference between discrimination and prescision becomes clear when we note that color can be discriminated, but not prescinded, from space: we can have an idea of color without having the idea of space, but we cannot suppose that there is any color not extended in space.

An effective way to see the difference between prescision and dissociation is to consider George Berkeley's critique of abstract ideas. In *A Treatise Concerning the Principles of Human Knowledge*, Berkeley points out that, whenever we try to imagine an abstract *man*, we can only imagine a man having a specific skin color, a specific stature, etc. We cannot imagine an abstract *man* or *humanity* without any specific skin color, stature, and so on, and therefore, he argues, we have no abstract idea of *man* (Berkeley [1710] 1998: Intro. §§9–10). To this Peirce might add: it is true that we cannot imagine a man without imagining him to have a specific skin color, specific stature, etc. However, it is possible to *neglect* the skin color, stature, etc. of a man and turn our *attention* to his humanity. In other words, although we cannot *dissociate* humanity from skin color, stature, and so on, we can *prescind* humanity from skin color, stature, etc. Dissociation is a psychological mode of separation based on our capacity to imagine certain states of affairs, whereas prescision is a logical mode of separation that is independent of our capacity of imagination.[43]

---

[42] A problem is that elsewhere Peirce writes that discrimination is reciprocal; e.g. in an earlier draft of the "New List" he writes: "If *A* can be discriminated or dissociated from *B*, *B* can also be separated from *A*, in the same mode" (W 1:519, 1866). Thus, if red cannot be discriminated from color, then color cannot be discriminated from red. Now if we assume that what cannot be discriminated can neither be prescinded nor dissociated (which seems to be implied by Peirce's statement that "[a]bstraction or prescision … supposes a greater separation than discrimination, but a less separation than dissociation"), then it follows that there is no way of mentally separating red and color, which is absurd. Perhaps Peirce ought to have said that red and color *can* be discriminated.

[43] It should be noted that Berkeley too seems to have in mind a mental operation similar to Peirce's prescision when he writes: "And here it must be acknowledged that a man may consider a figure merely as triangular, without attending to the particular qualities of the angles, or relations of the sides. So far he may abstract: but this will never prove, that he can frame an abstract general inconsistent idea of a triangle" (Berkeley [1710] 1998: Intro. §16). Where Peirce seems to differ from

§S.5 Method of deriving the categories (§6)

In §6, Peirce explains the method he will use to derive the categories intermediate between substance and being. The derivation procedure consists of two steps. The first step is to find conceptions that will be candidates for the categories. These conceptions, Peirce writes, are to be searched for in the data of "empirical psychology" (EP 1:3, W 2:51). The second step is to verify whether these conceptions actually qualify as categories. The method used in this verification procedure is the method of prescision outlined above. This second verifying step can be seen in the concluding sentence of each of the sections in which Peirce derives one of the intermediate categories (§§7–9), where he notes that the conception derived in that section cannot be prescinded from the preceding conception (in order of passing from being to substance; see below), but the latter can be prescinded from the former.

How does prescision function as a verification procedure? As Peirce writes in the second paragraph of §5, an important characteristic of prescision is that it is not a reciprocal process (EP 1:3, W 2:51). That is, it is often the case that, while a conception *A* cannot be prescinded from another conception *B*, *B* can be prescinded from *A*. Such a situation occurs because the prescindable conception *B* is *indispensable* in cognizing *A*, but once *A* has occasioned the introduction of *B*, it is generally possible to ignore *A*. For example, the manifold of sensuous impressions being united under the conception of space is a necessary condition for it to be united under the conception of color (hence color cannot be prescinded from space), but once the conception of space has been introduced, the color of the manifold, which occasioned the introduction of the conception of space, can generally be ignored (hence space can be prescinded from color). Generally speaking, if we are able to show that a conception *B* can be prescinded from another conception *A* but not vice versa, we have shown that *A* cannot be reduced to unity without the

---

Berkeley is that on Peirce's view, we *can* frame abstract and general (but presumably not "inconsistent") ideas by selective attention/prescision. These general ideas, however, are not *created* by the act of prescision; the ideas are already operative in the cognitive process, and prescision only isolates them from other ideas.

introduction of *B*. Now as we saw in §1, "the validity of a conception consists in the impossibility of reducing the content of consciousness to unity without the introduction of it." Therefore, if *B* can be prescinded from *A* but not vice versa, this means that *B* is valid in this sense (assuming that *A* is part of the "content of consciousness").

A category must be valid in this sense. Otherwise it would be superfluous, "a mere arbitrary addition" (EP 1:3, W 2:51), which would in turn make it non-universal and hence not a category. Therefore, given a conception *B* that we have already established to be a category, if we can find another conception *A* such that *B* can be prescinded from *A* but not vice versa, this means that *A* is a conception that makes *B* valid. Of course, this does not decisively prove that *A* is a category, since there might be other conceptions which also make *B* valid. But if there is such a conception, say *C*, then once it is found it can be verified using the same method. That is, given a conception *B* that we have already established to be a category, if *C* does not satisfy the condition that *B* can be prescinded from *C* but not vice versa, then *C* is thereby disqualified as a category.

Various views have been put forward by commentators regarding the method used by Peirce to derive the categories in the "New List." De Tienne, for example, argues that the method is inductive (De Tienne 1996:225). An important point to note in this connection is that Peirce's derivation of the categories does not proceed in the direction going from substance to being—which is the direction in which the cognitive process actually takes place—but rather in the opposite direction, *from being to substance*. This reversal of direction occurs toward the end of §6, where Peirce writes:

> Now, empirical psychology discovers the occasion of the introduction of a conception, and we have only to ascertain what conception already lies in the data which is united to that of substance by the first conception, but which cannot be supposed without this first conception, to have the next conception in order in passing *from being to substance*. (EP 1:3, W 2:51, emphasis added)

Why this reversal of direction? Probably because we never directly experience the starting point of the cognitive process, i.e., substance. We always find ourselves at being, the end point of the cognitive process, and so in order to uncover the structure of this process we must retrace our steps, going backward from being to substance.

According to De Tienne, this backward movement from being to substance is an inductive procedure. To see how it is supposed to work, let us consider the derivation of the first category in order of passing from being to substance, namely *quality*. De Tienne writes:

> [T]o find the conception that comes after that of being in the order of passage toward substance, it is enough to "observ[e] the occasion of the introduction of being" (W 1:520). This observation involves considering as many occurrences of the copula as possible (i.e., cases), and examining the kind of company that frequents it most regularly. As the observations are made, a certain regularity will eventually stand out and impose itself on the researcher, and will then be promoted to a conception by means of induction: Quality. (De Tienne 1996:226–27)

The same procedure can then be repeated to obtain the conceptions following that of quality, until we ultimately reach substance. Peirce's statement that the categories are to searched for in the data of "empirical psychology" reinforces this inductive reading of the derivation process.

Gabriele Gava (2011), on the other hand, sees Peirce's method of prescision as a transcendental method, where a *transcendental method* is understood as "an analysis of human experience and knowledge in general in order to abstract the fundamental elements, the conditions, without which such experience and knowledge would not be possible" (Gava 2011:235). Peirce's prescision is a transcendental method in this sense, because it shows which conceptions are indispensable in thinking others, and by applying it to experience and thought in general, he attempts to "to isolate those elements without which such experience and thought would have been unaccountable" (Gava 2011:236).

I believe both De Tienne and Gava are right. As we noted above, Peirce's derivation of the categories consists of two steps. The first step is to find in the data of empirical psychology the conceptions that will be candidates for the categories, while the second step is to verify whether

these conceptions qualify as categories using the method of prescision. The first step can be understood as an inductive procedure, in the sense maintained by De Tienne, while the second step can be understood as a transcendental procedure, in the sense maintained by Gava. It may also be fruitful to think of the first step as analogous to Kant's metaphysical deduction of the categories, and the second step to the transcendental deduction.

§S.6 Quality/reference to a ground (§7)

The first step we must take in deriving the first category in order of passing from being to substance is to find the conception that occasions the introduction of being. This, as we already saw above, is *quality*, which constitutes the topic of §7. Since the function of the conception of being is to unite a quality with the subject of attention (the substance) and thereby form a proposition, the conception of quality is the occasion for introducing that of being. As Peirce puts it in an earlier manuscript: "Character is the ground of being; whatever is, is by being *somehow*" (W 1:352, 1866).

As we saw in §2.2 of this dissertation, Peirce refers to quality as "reference to a ground." Since I have already discussed Peirce's notion of reference to a ground in §2.2, I will not repeat my comments here. The point to keep in mind is that every judgement, insofar as it is not made in a completely arbitrary way, must be made on the basis of some *ground*, which is a quality apprehended in itself, independently of its application to any specific circumstance.

§7 concludes with the following sentence: "Reference to a ground cannot be prescinded from being, but being can be prescinded from it" (EP 1:4, W 2:53). This, as we noted above, constitutes the second step of Peirce's derivation of the categories, in which he verifies that the derived conception (*quality* in this case) qualifies as a category by establishing that it validates the preceding category (*being* in this case).

§S.7 Relation/reference to a correlate (§8): preliminary discussion

In §8 Peirce takes up the second intermediate category in order of passing from being to substance, which he calls *relation* or *reference to a correlate*. This section, however, is brief and unhelpful:

> §8. Empirical psychology has established the fact that we can know a quality only by means of its contrast with or similarity to another. By contrast and agreement a thing is referred to a correlate, if this term may be used in a wider sense than usual. The occasion of the introduction of the conception of reference to a ground is the reference to a correlate, and this is, therefore, the next conception in order.
>
> Reference to a correlate cannot be prescinded from reference to a ground; but reference to a ground may be prescinded from reference to a correlate. (EP 1:5, W 2:53)

The brevity and obscurity of this section has led commentators to propose various interpretations as to what Peirce meant by *correlate*.[44] Here I want to take up the interpretation put forward by De Tienne in his book *L'analytique de la représentation chez Peirce* (De Tienne 1996), and offer my own interpretation by pointing out the problems with De Tienne's view.[45] Before going into De Tienne's interpretation, however, some preliminary clarifications are in order.

Peirce writes that "we can know a quality only by means of its contrast with or similarity to another." The first question that arises is: another *what*? Joseph Ransdell (1966:86) argues that Peirce intended the correlate to be a *form*—a quality, essence, or "firstness"—different from the quality constituted by reference to a ground. Against this De Tienne rightly points out: "If the correlate is a form or quality, then the relate should be one as well. But Peirce always speaks of the relate as substance-subject, that is, as something whose form is still undetermined" (De Tienne 1996: 287). Note that *relate* is a term that Peirce uses to refer to the initial substance (the sub-

[44] Apart from De Tienne (1996), which will be discussed below, see Murphey (1993, Chap. 3), Ransdell (1966:81–88), Michael (1980:198–201), and Ishida (2009:49–59).

[45] While my focus here will be on De Tienne's interpretation of the correlate, my critique of his view, I believe, will also carry over to other commentators who seem to assume that the relation between the relate and correlate must be one of similarity, such as Murphey (1993), or one of either similarity or dissimilarity, such as Michael (1980).

stance in need of explanation) in contexts where it is conceived as the first term of a (dyadic) relation. Contrary to Ransdell, the correlate should be thought of as a determinate thing or fact, as suggested by Peirce's statement (in a preliminary draft of the "New List") that "[a] correlate is a second substance with which the first is in comparison" (W 1:524, 1866).[46] It is the act of bringing the initial substance into relation with this second substance that occasions the introduction of reference to a ground, thereby making it possible to attribute a particular quality to the former.

§S.8 Relations of equiparance and disquiparance

Another source of confusion is that Peirce speaks of a "contrast" or "similarity" ("agreement") between the relate and correlate, which are quite different things. Here we should turn to §14, where Peirce draws a distinction between two kinds of relation:

> A quality may have a special determination which prevents its being prescinded from reference to a correlate. Hence there are two kinds of relation.
>
> 1st. That of relates whose reference to a ground is a prescindible or internal quality.
> 2nd. That of relates whose reference to a ground is an unprescindible or relative quality.
>
> In the former case, the relation is a mere *concurrence* of the correlates in one character, and the relate and correlate are not distinguished. In the latter case the correlate is set over against the relate, and there is in some sense an *opposition*. (EP 1:7, W 2:55)

Elsewhere, Peirce calls relations of the first kind *relations of equiparance* (or simply *equiparances*), and those of the second kind *relations of disquiparance* (or simply *disquiparances*) (e.g. W 1:475, 1866). This terminology is derived from medieval logic (see W 2:418–19fn12, 1870).

To understand this distinction, let us consider a concrete example: a weathercock. If we think of a situation where someone is looking at the weathercock, we can assume that the quality of

[46] See also Peirce's 1908 letter draft to Francis C. Russell, where he explains what he meant by correlate in the "New List": "What I call there [in the "New List"] a 'correlate' is an ordinary experiential correlate, reference to which is forced upon the mind. We may call it an 'occurrence,' meaning a *thing* or *fact*, single and definite" (RL 387b:329, 1908).

"indicating that the direction of the wind is such and such" is predicated of it. The weathercock itself is the relate, while the direction of the wind is the correlate of this predication. The above quality cannot be supposed without also supposing the correlate, the direction of the wind; that is, the quality cannot be prescinded from reference to the correlate. Therefore, the relation between the weathercock and the direction of the wind is a disquiparance. On the other hand, suppose that the weathercock has the same reddish-brown color as the roof it is attached to. The correlate of the weathercock's quality of being reddish-brown is the roof. In this case, the weathercock's quality of being reddish-brown can be supposed without also supposing the roof; that is, the quality can be prescinded from reference to the correlate. Therefore, the relation between the weathercock and the roof is an equiparance.

Peirce says that an equiparance is "a mere *concurrence* of the correlates in one character, and the relate and correlate are not distinguished." By this he means that interchanging the relate and correlate does not change the qualities ascribed to them; in other words the relation is symmetric. Thus, interchanging the weathercock and roof in the above example does not affect the predications "this weathercock is reddish-brown" and "this roof is reddish-brown." However, not all symmetric relations are equiparances; for example, the relation "____ equals ____" is symmetric but not an equiparance.[47] Since in an equiparance the reference to the correlate can be neglected, the relation effectively reduces to a monadic property; whereas a disquiparance is an irreducibly dyadic relation.

Peirce's statement in §8 that the relate and correlate are either in a relation of "contrast" or "similarity" ("agreement") should be understood as corresponding to this distinction between relations of equiparance and disquiparance. There is said to be a "similarity" ("agreement") between the relate and correlate when their relation is an equiparance, while they are said to be in

[47] In his writings of the 1860s, Peirce makes the error of identifying relations of equiparance (in the sense defined above) with symmetric relations and relations of disquiparance (in the sense defined above) with non-symmetric relations. This error is corrected in his "Description of a Notation for the Logic of Relatives" (W 2:418–19, 1870); see Michael (1974:64–68).

"contrast" when their relation is a disquiparance. With these clarifications in place, let us turn to De Tienne's interpretation of the correlate.

§S.9 De Tienne's interpretation of the correlate

According to De Tienne, in the case of relations of equiparance, "the correlate is a subject of past experience that has already undergone predication" (De Tienne 1996: 296–97). The idea here is that by comparing the relate—the substance in need of explanation—with the substance of another instance of predication that we know from past experience (the correlate), we are able to apply to the relate a quality similar or identical to the one that was applied to the correlate. This makes sense in the case of relations of equiparance, since in this case the relate and correlate are supposed to be similar or identical. However, a difficulty arises when we turn to relations of disquiparance.

Consider, for example, the relation of murder, which Peirce takes up in §9 in the context of explaining the notion of *representation*. He writes:

> Again, suppose we think of a murderer as being in relation to a murdered person; in this case we conceive the act of the murder, and in this conception it is represented that corresponding to every murderer (as well as to every murder) there is a murdered person; and thus we resort again to a mediating representation which represents the relate as standing for a correlate with which the mediating representation is itself in relation. (EP 1:5, W 2:53)

Here, Peirce is clearly thinking of the murderer as the relate and the murdered person as the correlate. De Tienne detects in this passage a certain confusion in Peirce's understanding of the correlate. He argues:

> Included in the conception of reference to a correlate is that of a reference to a treasure of past experiences—a stock of representations already carried out. Where does this reference appear in the relationship of the murderer to his victim? If the former is the relate, and the latter the correlate, we obviously cannot say that the reference to the victim is a reference to a past representation … But how can the interpretant put the murderer and victim into correlation if it cannot refer to an earlier representation, in which this correlation has already taken place? I think the main difference between "the stove is black" and "the murderer kills his victim" consists only in the number of subjects involved in

> the predicate, and that such a circumstance should not influence the general representation of the process as such. In other words, it should be possible to present in both cases a formally identical process. Thus, if the attribution of "black" to the stove is the function of a reference to a correlate that also underwent (or experienced) this attribution, this should also be the case with the attribution of the murder relationship between the murderer and his victim. (De Tienne 1996: 297)

The problem here is that Peirce seems to be using the term "correlate" in two incompatible senses, namely: (1) that which occasions the introduction of reference to a ground, and (2) the second term of a dyadic or triadic relation. In order to avoid this problem, De Tienne attempts a reformulation of the notion of correlate (De Tienne 1996: 297–98). He suggests that in the case of relations of disquiparance, the correlate should be understood not as the second term of the relation, but as an instance of the relation which one has already experienced in the past. In the case of the murder relation, the correlate is not the victim, but "an already determined image in which the relation of murderer to victim has already been established" (De Tienne 1996: 297–98). Thus, just as in the case of an equiparance, by comparing the unknown case at hand with an instance of murder that one has witnessed in the past (either in reality or in fiction), one is able to refer the case at hand to the same (or similar) ground that made sense of the past instance (i.e., reduced it to unity).

§S.10 My interpretation of the correlate

De Tienne assumes that the two senses in which Peirce uses the term "correlate"—namely, (1) that which occasions the introduction of reference to a ground, and (2) the second term of a dyadic or triadic relation—are incompatible. His reformulation consists in sacrificing the second sense in order to save the first; there would be no need for this if the two senses were compatible. I will argue that the two senses are indeed compatible. Hence, even in the case of dyadic relations of disquiparance (and also irreducibly triadic relations), we can understand the correlate in a way that is faithful to Peirce's texts: not as a past instance that is similar or identical to the case at hand, but as the second term of a dyadic or triadic relation. When the quality being

predicated of the relate is a monadic property, the correlate may be a past instance similar or identical to the case at hand; but even in this case it need not be so. For example, the correlate of the monadic predication "this is blue" may be a blue thing that one has experienced in the past, but it may also be the background with which the subject of predication is in contrast. In other words, on my view the correlate is a much broader notion than that of a past instance similar or identical to the case at hand.

I suggest that anything that can be regarded as being in relation with the initial substance should be understood as a *correlate*. This is indeed how Peirce uses the term; see e.g. Lecture IX of the 1866 Lowell Lectures, where he writes: "*Relate* and *correlate*, you remember, are terms employed to signify merely the thing related and the thing related *to*" (W 1:474). *It is by putting a thing into relation with other things that we gain knowledge about it*—reference to a correlate should be understood in this very broad sense.

How, then, are the two senses of the correlate compatible? A hint can be found in Peirce's 1894 rewriting of the "New List." The rewritten version of §8 opens as follows:

> The study of psychology, from which we find it convenient to borrow a few principles, shows us that we can never know, or even think, that a thing has a quality without thinking or having thought of other things partaking that quality and of still others wanting it, or at least possessing it in smaller measure. This is the natural, common-sense belief of the mass of men; and it seems to be confirmed by careful observation. There are only a few thinkers who do not accept it. This is the doctrine which ought in strictness to be called the *doctrine* of the *relativity of knowledge*. (R 403:11–12, 1894)

So far this seems to be in line with De Tienne's interpretation of the correlate, as a thing known from past experience partaking of a quality similar or identical to the one to be predicated in a judgment (although Peirce here also speaks of "others still wanting it"). But the question is not whether such a thing is a correlate; the question is whether this is the *only* kind of correlate, or, on the other hand, this is merely a special case. The passage that follows the above quotation is interesting:

> There is a corresponding truth in regard to existence. That is to say, things can only possess qualities by virtue of their mutual interactions. This proposition may be called the doctrine of the relativity of facts. For example, a thing cannot be *hard*, except by virtue of resisting other things; and if there were but one atom in the universe, to say that atom was hard would be a phrase without meaning. (R 403:12, 1894)

Consider the quality *hard*. Peirce is here claiming that the hardness of an object consists in the resistance of that object against other objects. For example, to say that a diamond is hard is to say that it would not be scratched if one were to apply pressure to it with, say, a knife-edge. The connection with Peirce's pragmatism is evident. In fact, the example of *hard* is the same example that Peirce gives to illustrate his pragmatic maxim in "How to Make Our Ideas Clear" (EP 1:132, W 3:266, 1878). But it is important to realize the context in which this discussion is taking place: namely, the rewriting of §8 of the "New List," whose topic is reference to a correlate. Peirce's point is clearly that the other object—the object being resisted by the hard object—is the correlate of the predication "this is hard." In the above example, the diamond is the relate and the knife-edge is the correlate. Here we can see how a reference to a correlate, where the correlate is understood as the second term of a relation (in this case the relation of resistance), can occasion the attribution of a quality to the relate. It is by observing that a diamond resists the pressure of a knife-edge that induces us to attribute the quality *hard* to it. Of course, an object need not be constantly in resistance against something in order for us to say that it is hard. Hence, the quality *hard* can be prescinded from reference to the object being resisted, which is what makes *hard* an internal (monadic) quality. The relation of resistance, on the other hand, is a disquiparance—one cannot suppose that there is a resistance without referring to the object being resisted.

Note that I am not trying to make the anachronistic claim that Peirce's 1894 rewriting of the "New List" faithfully represents what Peirce had in mind in 1867. Rather, my intention in citing the 1894 version is simply to illustrate how the two senses of the correlate can be understood as being compatible, even in the case of relations of disquiparance.

§S.11 The notion of comparison

In §9, Peirce argues that "[t]he occasion of reference to a correlate is obviously by comparison" (EP 1:5, W 2:53), and goes into a discussion of the *comparison* of the relate and correlate. The problem with De Tienne's view of the correlate is that it is based on a narrow, psychological reading of Peirce's notion of comparison. That is, on his reading the comparison of a relate and correlate is restricted to the act of observing their similarity or dissimilarity with respect to a certain quality. This, as I see it, is what leads him to restrict the correlate to a past instance that is similar or identical to the case at hand.

For Peirce, however, the term "comparison" has a much broader meaning than this psychological sense. This can be seen not only in his discussion of "comparing" a murderer with the victim in §9, but also more explicitly in the following manuscript from 1865:

> [E]verything is such as it is in comparison with something else. This is an old and established axiom … the effect of this ancient maxim is that '*blue*' MEANS 'blue in comparison to' and therefore requires a suffering object [i.e., correlate]. The transitive verb supplies this comparison. If a man kills a deer, that in comparison to which he is a killer is the deer. No other comparison is *needed*. (W 1:336, 1865)

From this it is clear that Peirce's notion of comparison is much broader than the psychological sense of observing a similarity or dissimilarity in quality. It follows that there is no need to restrict the correlate to a similar or identical past representation in order to make sense of Peirce's discussion of "comparing" the relate and correlate.

To sum up, the problem with De Tienne's interpretation of the correlate is that by conceiving it in a narrow, psychological sense, it can no longer be understood as the second term of an arbitrary relation, contrary to what Peirce says. On the other hand, the broader conception of the correlate that I am proposing, according to which a correlate is anything that can be regarded as being in relation with the initial substance, is not only faithful to Peirce's texts, but also has the further advantage that it allows us to see the latent connection between Peirce's notion of the correlate and his later pragmatism, as displayed in the example of *hardness* that we saw above.

§S.12 Representation/reference to an interpretant (§9)

In §9 Peirce takes up the third category in order of passing from being to substance, which he calls "representation" or "reference to an interpretant." In his later writings Peirce will also call the interpretant a "Third." Before going into the text of the "New List," let me quote a few definitions of a "Third" from Peirce's later writings:

> The Third is that which is what it is owing to things between which it mediates and which it brings into relation to each other. (EP 1:248, 1887–88)
>
> [A Third is that] whose Being consists in active power to establish connections between different objects … (EP 2:435, 1908)

Just like the Third described in these passages, the interpretant discussed in §9 is something which, given a relate, calls forth a correlate and establishes a relation between them. It is the reference to such an interpretant that occasions the introduction of reference to a correlate. We will look at some specific examples below. As we noted above, Peirce begins §9 with the statement that "[t]he occasion of reference to a correlate is obviously by comparison" (EP 1:5, W 2:53). Here, it is important to keep in mind that the term "comparison" does not necessarily signify a comparison with respect to a certain quality. As I have already argued, what Peirce calls a "correlate" is much broader than a similar or identical past representation, and hence, the relation established between the relate and correlate need not be one of similarity or identity—it can be any kind of relation.

Peirce gives three examples to explain the process of comparison. The first example is a comparison of the letters "p" and "b." Let us assume that "p" is the relate and "b" is the correlate (the choice is arbitrary; the same explanation would hold even if we interchanged the roles of the letters). One way to compare these letters is to imagine that "b" is rotated with the line of writing as an axis, laid over "p," and made transparent so that "p" can be seen through it. The "image" of this series of operations is the interpretant, because this image is what brings the two letters into relation (Fig. A.1).

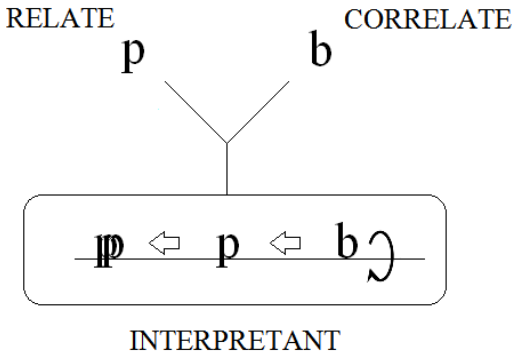

Fig. S.1. Comparing the letters “p” and “b.”

Peirce writes that this image “mediates between the images of the two letters, inasmuch as it represents one of them to be (when turned over) the likeness of the other” (EP 1:5, W 2:53). Let me briefly explain Peirce’s use of the term “represent” and its noun form, “representation.” Peirce often uses the term “represent” in explaining the function of signs. While the word “sign” does not appear in §9 of the “New List,” it is clear that he has the sign relation in mind in this section. According to Peirce’s later theory of signs, a sign consists of three elements: the *representamen* or sign itself, the sign’s *object*, and the *interpretant* which connects these two. In the “New List,” the relate corresponds to the representamen, and the correlate corresponds to the object. Now an essential characteristic of signs is that, even if an object is not actually present, a sign has the power to substitute for the object for anyone who recognizes the sign as such. For example, whenever we want to speak with someone about a certain book, we can refer to the book by uttering its title, without having to actually carry it around and pointing at it with our finger. It is in this sense that Peirce calls a sign a “representation”—it is a *re*-presentation of the object it signifies.

In the above example, the interpretant—the image of the series of operations of flipping one of the letters, laying it over the other, and making it transparent—is said to represent one of the letters to be the likeness of the other. That is, “p” is able to function as a representation of “b” by virtue of the fact that the interpretant represents “p” to be a representation of “b.” Thus, the

interpretant itself is also a sign that represents "p" as representing "b". We will touch upon this point again below.

Let us return to Peirce's examples of comparison. The second example that he takes up is that of murder, which we have already mentioned in connection with the correlate (§A.9). Suppose that the murdered person is the relate and the murderer is the correlate (for the sake of illustration, I have exchanged the role of the relate and correlate from Peirce's example). Let us imagine a specific situation where there is a dead body lying on the roadside with a knife stabbed in its chest. The first thing that will come to mind when we witness this scene is most likely the idea of murder. And since corresponding to every murdered person (and every act of murder) there must be a murderer, we infer that there must also be a murderer who killed the particular victim in front of us. Thus, the concept of murder calls forth a correlate, the murderer, and brings it into relation with the relate, the dead body. In this case the dead body is functioning as a sign of the murderer. The interpretant here is the concept of murder possessed by whoever finds the dead body.

The third example is that of looking up the word "homme" in a French-English dictionary. When we do so, we find next to it the word "man." Peirce writes that "the word *man* … so placed, represents *homme* as representing the same two-legged creature which *man* itself represents" (EP 1:5, W 2:53). In this case, the word "homme" is the relate (sign), the two-legged creature which it represents is the correlate (object), and the word "man" placed next to the word "homme" in a French-English dictionary is the interpretant (Fig. A.2). Note that the word "man" does not function as an interpretant by itself. The word "man" is able to function as an interpretant only when it is placed in an appropriate context, in this case, next to the word "homme" in a French-English dictionary.

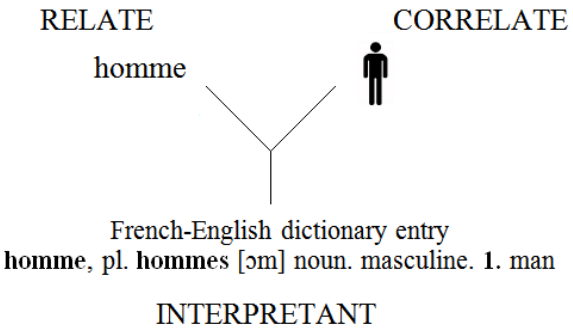

Fig. S.2 Looking up the word "homme" in a French-English dictionary

On the basis of these examples (and an "accumulation" of further unspecified instances), Peirce defines an interpretant as follows: "*a mediating representation which represents the relate to be a representation of the same correlate which this mediating representation itself represents*" (EP 1:5, W 2:53, emphasis in original). This representation is called an *interpretant* "because it fulfils the office of an interpreter, who says that a foreigner says the same thing which he himself says" (EP 1:5, W 2:54). Let us examine the analogy of the interpreter in more detail in order to analyze what Peirce is saying in his definition of the interpretant.

§S.13 The double function of the interpretant

Joseph Ransdell (1966) pays special attention to Peirce's interpretant/interpreter analogy. Building on this analogy, he considers an interpreting situation where one man, *A*, speaks in some language, and another man, *B*, repeats what *A* says in a different language (Ransdell 1966: 74). What is it that makes *B* an interpreter in this situation? It is not the content of his utterances as such, but rather the role or position that he occupies in the given context. In Peirce's description of the interpreter as someone "who says that a foreigner says the same thing which he himself says," the contextual role is expressed by the clause "*who says that*." The interpreter may make his role explicit by prefixing all of his utterances with a clause such as "*A* says that …," but regardless of whether the interpreter actually says this, "it is implicitly understood that he is

saying this—for otherwise he would not be functioning as interpreter. Hence, the interpreter, as such, always represents himself to be such" (Ransdell 1966: 75). That is, the interpreter does not only provide a correlation between the foreign utterances and their meaning; he also presents himself *as* an interpreter. These are two distinct roles played by the interpreter.

The same can be said of the interpretant. The first function of the interpretant is, given a relate (sign), to call forth a correlate (object) and present the relate as a representation of the correlate, thereby bringing them into correlation. The second function of the interpretant is to present itself to a second interpretant (the audience in the above example) as representing, in place of the relate, the same correlate which the relate represents. Thus, in its second function the interpretant itself is also a sign which presents itself to further interpretants. This process may in principle continue indefinitely, and is sometimes referred to as "infinite semiosis" by semioticians and Peirce scholars.[48]

§S.14 The reversal of Kant (§10)

§10 is intended to show that reference to an interpretant is the last category in order of passing from being to substance, and hence there are a total of three intermediate categories. Peirce claims that the occasion for reference to an interpretant is the diversity of sense impressions:

> If we had but one impression, it would not require to be reduced to unity, and would therefore not need to be thought of as referred to an interpretant, and the conception of reference to an interpretant would not arise. But since there is a manifold of impressions, we have a feeling of complication or confusion, which leads us to differentiate this impression from that, and then, having been differentiated, they require to be brought to unity. (EP 1:6, W 2:54)

The sentence that follows is important: "Now they [the impressions] are not brought to unity until we conceive them together as being *ours*, that is, until we refer them to a conception as their interpretant" (EP 1:6, W 2:54). That is, to say that the impressions are brought to unity

[48] See Aames (2018) for a more detailed discussion of the two roles played by the interpretant in Peirce's theory of signs.

means that they are conceived together as being *ours*, and this is made possible by referring them to an interpretant. For example, consider a situation where we suddenly remember the melody of a musical piece that we heard somewhere long ago, but we cannot recall what the musical piece is or where we heard it. In this case, the melody is the manifold of impressions, and the fact that we are trying to remember what the musical piece is corresponds to the manifold calling to be brought to unity. It is the role of an interpretant (or series of interpretants) to bring this manifold to unity by connecting it to such correlates as the place where we heard the musical piece and the rest of the melody. Through the function of the interpretant, the forgotten melody once again becomes a part of us and is conceived as being "our" impression.

As De Tienne (1996:316–17) points out, we can discern in this section a reversal of Kant's idea of the transcendental apperception. Kant held that experience in general is made possible by what he called transcendental apperception, which is the function of unifying all appearances into one consciousness by referring them to a common subject, by attributing them to the self as "my" experience: "The *I think* (*Ich denke*) must be *capable* of accompanying all my presentations. For otherwise something would be presented to me that could not be thought at all—which is equivalent to saying that the presentation either would be impossible, or at least would be nothing to me" (B 132).[49] Contrary to this, Peirce is claiming that it is the categories' function of reducing the manifold of sense impressions to unity that makes it possible to conceive of impressions as belonging to "us," and thereby bring about the unity of consciousness. As De Tienne puts it: "Consciousness is one by virtue of the unity proper to representation, and not the reverse: it is not the representation that is one because it is the fact of a transcendental 'I think'" (De Tienne 1996:316–17).

[49] Translation by Werner S. Pluhar (Kant 1996:177).

§S.15 The list of categories (§§11–13)

The argument so far affords us with the following list of categories, which Peirce presents in §11:

Being
    Quality (Reference to a Ground)
    Relation (Reference to a Correlate)
    Representation (Reference to an Interpretant)
Substance

Peirce calls the three intermediate categories "accidents." In traditional Aristotelian terminology, the term "accident" is used in contrast to "essence." Roughly, a property is said to be an "essential" property of something if it is necessarily involved in that thing, whereas a property is said to be "accidental" if it is not necessarily involved in that thing. Thus, humanity is an essential property of Socrates, whereas his being seated is an accidental property. Peirce calls the three intermediate categories "accidents" probably because quality, relation, and representation can take various specific forms, while "to be" and "to be present" are properties necessarily involved in everything.

Next, in §12 Peirce describes the function of the categories in numerical terms. The entire section is as follows:

> §12. This passage from the many to the one is numerical. The conception of a *third* is that of an object which is so related to two others, that one of these must be related to the other in the same way in which the third is related to that other. Now this coincides with the conception of an interpretant. An *other* is plainly equivalent to a *correlate*. The conception of second differs from that of other, in implying the possibility of a third. In the same way, the conception of *self* implies the possibility of an *other*. The *Ground* is the self abstracted from the concreteness which implies the possibility of an other. (EP 1:6, W 2:55).

The initially given substance is a "manifold," and the reduction of this manifold to unity proceeds in the order of three, two, one. While this movement from the many to one is a Kantian motif, the ordinal reinterpretation of an "other" as "second," and the idea that the "conception of

second differs from that of other, in implying the possibility of a third" is a uniquely Peircean view. The same can be said of the claim that "the conception of *self* implies the possibility of an *other*." This ordinal conception of the categories prefigures the terminology of Peirce's later theory, where the three intermediate categories are simply called Firstness, Secondness, and Thirdness (though the term "first" does not appear in the "New List").

Finally, in §13 Peirce presents the following list of "supposable objects" afforded by the categories (EP 1:6, W 2:55):

What is
    Quale (that which refers to a ground)
    Relate (that which refers to ground and correlate)
    Representamen (that which refers to ground, correlate, and interpretant)
It

What we should notice here is that, unlike "Relation" in §11, "Relate" is characterized as "that which refers to ground and correlate," and similarly, unlike "Representation" in §11, "Representamen" is characterized as "that which refers to ground, correlate, and interpretant." In contrast to the list of §11, which is a list of categories, the list in this section is a list of the objects that can be supposed by prescinding some of the categories that have been joined to substance. The objects lower down this list have a richer content (refer to more things) because a category cannot be prescinded from those above it in the list of §11.

Thus, supposing that the initial substance has been reduced to the unity of being, if we do not prescind away any of the categories, then we have a "Representamen," which is the substance with all three intermediate categories adjoined to it. If we prescind the reference to a correlate from reference to an interpretant, and thereby ignore the reference to an interpretant, we have a "Relate," which is the substance together with reference to a ground and reference to a correlate. Similarly, if we prescind the reference to a ground from reference to a correlate, and thereby ignore the reference to a correlate, we have a "Quale," which is the substance together with ref-

erence to a correlate. Finally, if we prescind the reference to an interpretant from substance, we have a bare "It," and if we prescind being from reference to a ground, we simply have "What is."

§S.16 Regularity as the basis of cognizability

So far I have given an outline of Peirce's theory of cognition presented in the "New List." How does this tie together with the theme of patternhood that we have been dealing with in this dissertation? The crucial observation is that in Peirce's categorial scheme, generals, including patterns, are Thirds: their being "consists in active power to establish connections between different objects" (EP 2:435, 1908).

This is closely related to Peirce's pragmatism. As we saw in §2.4, according to the pragmatic maxim, to predicate a general of a particular object x is to judge that x is governed by a series of laws or regularities that dictate how x would behave in certain hypothetical situations. These laws or regularities constitute the *intellectual purport* of the general in question—roughly, that part of the general that has a power of influencing actions and events, as distinguished from the purely qualitative aspect of it. Now a law or regularity is something that establishes connections between different objects. For example, if we find a dead body on the roadside with a knife stabbed in its chest, we are led by habit—which is a kind of regularity—to think of a murderer. The habit establishes a connection between the dead body and murderer. The concept of murder is an embodiment of this and other habits, which together constitute the intellectual purport of the concept. It is in this sense that a general is said to be a Third.

As we have seen so far in this chapter, according to Peirce's theory of cognition presented in the "New List," reference to an interpretant is a category, a necessary element in any act of cognition whatsoever. The interpretant corresponds to the "intellectual purport" of the predicate being applied to the substance, and as such, its mode of being is that of a law or regularity.[50]

[50] In a 1907 essay, Peirce makes a distinction between three types of interpretant: the *emotional interpretant*, *energetic interpretant*, and *logical interpretant* (CP 5.475–76, 1907). It should be noted

Therefore, as I suggested without argument in the introduction to this dissertation, we can say that regularity is a necessary condition for the very possibility of cognition.

To cognize something is to treat it as a node within a network of patterns and regularities. This is illustrated vividly by Peirce's pragmatic clarification of *lithium* from the third section of his "A Syllabus of Certain Topics of Logic," composed to accompany his 1903 Lowell Lectures:

> If you look into a textbook of chemistry for a definition of lithium, you may be told that it is that element whose atomic weight is 7 very nearly. But if the author has a more logical mind he will tell you that if you search among minerals that are vitreous, translucent, grey or white, very hard, brittle, and insoluble, for one which imparts a crimson tinge to an unluminous flame, this mineral being triturated with lime or witherite rats-bane, and then fused, can be partly dissolved in muriatic acid; and if this solution be evaporated, and the residue be extracted with sulphuric acid, and duly purified, it can be converted by ordinary methods into a chloride, which being obtained in the solid state, fused, and electrolyzed with half a dozen powerful cells, will yield a globule of a pinkish silvery metal that will float on gasolene; and the material of that is a specimen of lithium. (EP 2:286, 1903)

Here, we can see that the meaning of the concept of *lithium* is clarified by referring it to a series of patterns and regularities: lithium is a metal extracted from a mineral that "imparts a crimson tinge to an unluminous flame," "can be partly dissolved in muriatic acid" after being "triturated with lime or witherite rats-bane, and then fused," and so on. It is true that Peirce also mentions properties such as "vitreous" and "translucent" in his definition of lithium, but these in turn can be further clarified into a series of regularities via the pragmatic maxim. The central point I want to make here is that to cognize something, to render something intelligible, is to situate it within a network of patterns and regularities involving that thing, as seen in the example of lithium above. Hence, a thing that does not exhibit any kind of patternhood or regularity which the mind can seize upon would *ipso facto* not be a possible object of cognition.

---

that among these, only the logical interpretant can be regarded as a law or regularity; the emotional and energetic interpretants do not have the mode of being of a law or regularity.

Furthermore, as we saw in §2.3, according to Peirce's basic idealism "*cognizability* (in its widest sense) and *being* are not merely metaphysically the same, but are synonymous terms" (EP 1:25, W 2:208–9, 1868). From this it follows that a thing that does not exhibit any kind of patternhood or regularity would not have any being at all. We are thus led back to the passage quoted as the epigraph to the first chapter of this dissertation: "Generality is, indeed, an indispensable ingredient of reality; for mere individual existence or actuality without any regularity whatever is a nullity. Chaos is pure nothing" (EP 2:343, 1905).

Abbreviations

CP x.y = *Collected Papers of Charles Sanders Peirce*, volume x, paragraph y.

W x:y = *Writings of Charles S. Peirce: A Chronological Edition*, volume x, page y.

RLT x = *Reasoning and the Logic of Things: The Cambridge Conferences Lectures of 1898*, page x.

EP x:y = *The Essential Peirce: Selected Philosophical Writings*, volume x, page y.

PM x = *Philosophy of Mathematics: Selected Writings*, page x.

ILS x = *Illustrations of the Logic of Science*, page x.

R x:y = Manuscript housed in Harvard University's Houghton Library. The number x signifies the catalogue number assigned by Richard S. Robin in his *Annotated Catalogue of the Papers of Charles S. Peirce*; y is the sheet number.

RL x: y = Correspondence housed in Harvard University's Houghton Library. The number x signifies the catalogue number assigned by Richard S. Robin in his *Annotated Catalogue of the Papers of Charles S. Peirce*; y is the sheet number.

Bibliography

Aames, Jimmy J. 2016. *Patternhood, Correlation, and Generality: Foundations of a Peircean Theory of Patterns*. MA Thesis (Indiana University). Available at: https://scholarworks.iupui.edu/handle/1805/10896 (Accessed May 4, 2020)

———. 2018. "The Double Function of the Interpretant in Peirce's Theory of Signs." *Semiotica* 2018 (225): 39–55.

Adams, Marilyn McCord. 1982. "Universals in the Early Fourteenth Century," in *The Cambridge History of Later Medieval Philosophy: From the Rediscovery of Aristotle to the Disintegration of Scholasticism, 1100–1600*, eds. Norman Kretzmann, Anthony Kenny, & Jan Pinborg. Cambridge: Cambridge University Press.

Alexander, Samuel. 1920. *Space, Time, and Deity: The Gifford Lectures at Glasgow, 1916–1918*, 2 vols. London: Macmillan.

Anderson, Philip W. 1972. "More is Different: Broken Symmetry and the Nature of the Hierarchical Structure of Science." *Science* 177 (4047): 393–96.

Bedau, Mark A. 1997. "Weak Emergence." *Noûs* 31 (s11): 375–99.

———. 2002. "Downward Causation and the Autonomy of Weak Emergence." *Principia* 6 (1): 5–50.

Bennett, Charles H. 1988. “Logical Depth and Physical Complexity,” in *The Universal Turing Machine: A Half-Century Survey*, ed. Rolf Herken. Oxford: Oxford University Press.

Berkeley, George. [1710] 1998. *A Treatise Concerning the Principles of Human Knowledge*, ed. Jonathan Dancy. Oxford: Oxford University Press.

Braithwaite, Richard Bevan. 1953. *Scientific Explanation: A Study of the Function of Theory, Probability and Law in Science*. Cambridge: Cambridge University Press.

Buzzelli, Donald E. 1972. “The Argument of Peirce’s ‘New List of Categories.’” *Transactions of the Charles S. Peirce Society* 8 (2): 63–89.

Chaitin, Gregory J. 1975. “Randomness and Mathematical Proof.” *Scientific American* 232 (5): 47–52.

Chalmers, David J. 2006. “Strong and Weak Emergence,” in *The Re-Emergence of Emergence: The Emergentist Hypothesis From Science to Religion*, eds. Philip Clayton & Paul Davies. Oxford: Oxford University Press.

Dennett, Daniel C. 1971. “Intentional Systems.” *The Journal of Philosophy* 68 (4): 87–106.

———. 1987. “True Believers,” in *The Intentional Stance*. Cambridge, MA: MIT Press.

———. 1991. “Real Patterns.” *The Journal of Philosophy* 88 (1): 27–51.

———. 2000. “With a Little Help from My Friends,” in *Dennett’s Philosophy: A Comprehensive Assessment*, eds. Don Ross, Andrew Brook, & David Thompson. Cambridge, MA: MIT Press.

De Tienne, André. 1996. *L’analytique de la représentation chez Peirce: La genèse de la théorie des catégories*. Bruxelles: Facultes universitaires Saint-Louis.

Dewan, E. M. 1976. “Consciousness as an Emergent Causal Agent in the Context of Control System Theory,” in *Consciousness and the Brain: A Scientific and Philosophical Inquiry*, eds. Gordon G. Globus, Grover Maxwell, & Irwin Savodnik. Boston: Springer US.

Emmeche, Claus, Simo Køppe, & Frederik Stjernfelt. 2000. “Levels, Emergence and Three Versions of Downward Causation,” in *Downward Causation: Minds, Bodies and Matter*, eds. Peter Bøgh Andersen, Claus Emmeche, Niels Ole Finnemann, & Peder Voetmann Christiansen. Aarhus: Aarhus University Press.

Fisch, Max H. [1967] 1986. “Peirce’s Progress From Nominalism Toward Realism,” in *Peirce, Semeiotic, and Pragmatism: Essays by Max H. Fisch*, eds. Kenneth L. Ketner & Christian J. W. Kloesel. Bloomington, IN: Indiana University Press.

Gava, Gabriele. 2011. “Peirce’s ‘Prescision’ as a Transcendental Method.” *International Journal of Philosophical Studies* 19 (2): 231–53.

Hoel, Erik P. 2017. “When the Map Is Better Than the Territory.” *Entropy* 19 (5): 188.

———. 2018. “Agent Above, Atom Below: How Agents Causally Emerge from Their Underlying Microphysics,” in *Wandering Towards a Goal: How Can Mindless Mathematical Laws Give Rise to Aims and Intention?* eds. Anthony Aguirre, Brendan Foster, & Zeeya Merali. Springer.

Hulswit, Menno. 2005. “How Causal is Downward Causation?” *Journal for General Philosophy of Science / Zeitschrift für allgemeine Wissenschaftstheorie* 36 (2): 261–87.

Humphreys, Paul. 2008. “Synchronic and Diachronic Emergence.” *Minds and Machines* 18: 431–42.

———. 2016. *Emergence: A Philosophical Account*. Oxford: Oxford University Press.

Huygens, Christiaan. 1893. *Œuvres complètes de Christiaan Huygens, Tome V: Correspondance 1664–1665*. Den Haag: Martinus Nijhoff.

Ishida, Masato. 2009. *A Philosophical Commentary on C. S. Peirce’s “On A New List of Categories”: Exhibiting Logical Structure and Abiding Relevance*. Doctoral Dissertation (Pennsylvania State University). Retrieved from: http://etda.libraries.psu.edu/paper/9857/4344 (Accessed June 15, 2020)

———. 2013. “A Peircean Reply to Quine’s Two Problems.” *Transactions of the Charles S. Peirce Society* 49 (3): 322–47.

James, William. 1987. *William James: Writings 1902–1910*. New York: Library of America.

Kaneko, Kunihiko. 2006. *Life: An Introduction to Complex Systems Biology*. Springer.

Kant, Immanuel. 1996. *Critique of Pure Reason*, trans. Werner S. Pluhar. Indianapolis, IN: Hackett.

Kemling, Jared. 2018. “Peirce’s Transcendental Method: The Latent Debate between Prescision and Abduction.” *Transactions of the Charles S. Peirce Society* 54 (2): 249–72.

Kim, Jaegwon. 1992. “‘Downward Causation’ in Emergentism and Nonreductive Physicalism,” in *Emergence or Reduction?: Essays on the Prospects of Nonreductive Physicalism*, eds. Ansgar Beckermann, Hans Flor, & Jaegwon Kim. Berlin: De Gruyter.

———. 1999. “Making Sense of Emergence.” *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition* 95(1/2): 3–36.

———. 2000. “Making Sense of Downward Causation,” in *Downward Causation: Minds, Bodies and Matter*, eds. Peter Bøgh Andersen, Claus Emmeche, Niels Ole Finnemann, & Peder Voetmann Christiansen. Aarhus: Aarhus University Press.

Ladyman, James. 1998. “What is Structural Realism?” *Studies in History and Philosophy of Science Part A* 29 (3): 409–24.

Ladyman, James, Don Ross, et al. 2007. *Every Thing Must Go: Metaphysics Naturalized*. Oxford: Oxford University Press.

Lane, Robert. 2018. *Peirce on Realism and Idealism*. Cambridge: Cambridge University Press.

Locke, John. [1689] 1975. *An Essay Concerning Human Understanding*, ed. Peter H. Nidditch. Oxford: Oxford University Press.

Mayr, Ernst. 1974. "Teleological and Teleonomic, a New Analysis," in *Methodological and Historical Essays in the Natural and Social Sciences*, eds. Robert S. Cohen & Marx W. Wartofsky. Springer.

Michael, Emily. 1974. "Peirce's Early Study of the Logic of Relations, 1865–1867." *Transactions of the Charles S. Peirce Society* 10 (2): 63–75.

Michael, Fred. 1980. "The Deduction of the Categories in Peirce's 'New List.'" *Transactions of the Charles S. Peirce Society* 16 (3): 179–211.

Misak, C. J. 2004. *Truth and the End of Inquiry: A Peircean Account of Truth*, Expanded Paperback Edition. Oxford: Oxford University Press.

Murphey, Murray G. 1965. "On Peirce's Metaphysics." *Transactions of the Charles S. Peirce Society* 1 (1): 12–25.

———. 1993. *The Development of Peirce's Philosophy*, Revised Edition. Indianapolis, IN: Hackett.

Pasachoff, Jay M. & Alex Filippenko. 2019. *The Cosmos: Astronomy in the New Millennium*, 5th Edition. Cambridge: Cambridge University Press.

Peirce, Charles S. 1931–1958. *Collected Papers of Charles Sanders Peirce*, vols. 1–6, eds. Charles Hartshorne & Paul Weiss (1931–1935); vols. 7 & 8, ed. Arthur W. Burks (1958). Cambridge, MA: Harvard University Press.

———. 1982–2009. *Writings of Charles S. Peirce: A Chronological Edition*, ed. Peirce Edition Project. Bloomington, IN: Indiana University Press.

———. 1992. *Reasoning and the Logic of Things: The Cambridge Conferences Lectures of 1898*, ed. Kenneth L. Ketner. Cambridge, MA: Harvard University Press.

———. 1992–1998. *The Essential Peirce: Selected Philosophical Writings*, vol. 1, eds. Nathan Houser & Christian Kloesel (1992); vol. 2, ed. Peirce Edition Project (1998). Bloomington, IN: Indiana University Press.

———. 2011. *Philosophy of Mathematics: Selected Writings*, ed. Matthew E. Moore. Bloomington, IN: Indiana University Press.

———. 2014. *Illustrations of the Logic of Science*, ed. Cornelis de Waal. Chicago: Open Court.

Pikovsky, Arkady, Michael Rosenblum, & Jürgen Kurths. 2001. *Synchronization: A Universal Concept in Nonlinear Sciences*. Cambridge: Cambridge University Press.

Putnam, Hilary. 1975. *Mind, Language and Reality: Philosophical Papers, Vol. 2*. Cambridge, MA: Harvard University Press.

Quine, Willard van Orman. 2013. *Word and Object*, New Edition. Cambridge, MA: MIT Press.

Ransdell, Joseph M. 1966. *Charles Peirce: The Idea of Representation.* Doctoral Dissertation (Columbia University). Retrieved from: http://www.iupui.edu/~arisbe/rsources/dissabs/ransdell.htm (Accessed June 15, 2020)

Roberts, Don. D. 1970. "On Peirce's Realism." *Transactions of the Charles S. Peirce Society* 6 (2): 67–83.

Robin, Richard S. 1967. *Annotated Catalogue of the Papers of Charles S. Peirce*. Cambridge, MA: University of Massachusetts Press.

Ross, Don. 2000. "Rainforest Realism: A Dennettian Theory of Existence," in *Dennett's Philosophy: A Comprehensive Assessment*, eds. Don Ross, Andrew Brook, & David Thompson. Cambridge, MA: MIT Press.

Seager, William. 2012. *Natural Fabrications: Science, Emergence and Consciousness*. Springer.

Short, T. L. 2013. "Questions Concerning Certain Claims Made for the 'New List.'" *Transactions of the Charles S. Peirce Society* 49 (3): 267–98.

Sperry, R. W. 1969. "A Modified Concept of Consciousness." *Psychological Review* 76 (6): 532–36.

Strawson, Galen. 2006. "Realistic Monsim: Why Physicalism Entails Panpsychism," in *Consciousness and Its Place in Nature: Does Physicalism Entail Panpsychism?* ed. Anthony Freeman. Exeter: Imprint Academic.

Strogatz, Steven H. 2015. *Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering*, 2nd Edition. Boulder, CO: Westview Press.

Sugano, Reiji. 2013. *Fukuzatsukei Kagaku no Tetsugaku Gairon* [An Introduction to the Philosophy of Complex Systems Science]. Tokyo: Hon no Izumi.

Syropoulos, Apostolos. 2008. *Hypercomputation: Computing Beyond the Church-Turing Barrier*. Springer.

Weinberg, Steven. 1987. "Newtonianism, Reductionism and the Art of Congressional Testimony." *Nature* 330: 433–37.

Wiener, Norbert. 1965. *Cybernetics: or Control and Communication in the Animal and the Machine*, 2nd Edition, Cambridge, MA: MIT Press.

Wimsatt, William C. 1976. “Reductionism, Levels of Organization, and the Mind-Body Problem,” in *Consciousness and the Brain: A Scientific and Philosophical Inquiry*, eds. Gordon G. Globus, Grover Maxwell, & Irwin Savodnik. Boston: Springer US.

———. 1994. “The Ontology of Complex Systems: Levels of Organization, Perspectives, and Causal Thickets.” *Canadian Journal of Philosophy* 24 (sup1): 207–74.

Wolfram, Stephen. 1985. “Undecidability and Intractability in Theoretical Physics.” *Physical Review Letters* 54 (8): 735–38.