Computer Modeling in Philosophy

Arthur C. Schwaninger*

What Simulations Teach Us About Ordinary Objects

Abstract: Under the label of *scientific metaphysics*, many naturalist metaphysicians are moving away from *a priori* conceptual analysis and instead seek scientific explanations that will help bring forward a unified understanding of the world. This paper first reviews how our classical assumptions about ordinary objects fail to be true in light of quantum mechanics. The paper then explores how our experiences of ordinary objects arise by reflecting on how our neural system operates algorithmically. Contemporary models and simulations in computational neuroscience are shown to provide a theoretical framework that does not conflict with existing fundamental physical theories, and nonetheless helps us make sense of the manifest image. It is argued that we must largely explain how the manifest image arises in algorithmic terms, so that we can pursue a metaphysics about ordinary objects that is scientifically well founded.

Keywords: scientific metaphysics, eliminativism, ordinary objects, cognitive metaphysics, generative networks, manifest image, predictive processing

1 Introduction

There seems to be an everlasting battle in metaphysics regarding how to deal with ordinary objects. Common sense intuition supports the conservative view that things are as they seem: there exist tables, chairs, houses, and trees. However, revisionists view the world differently and have presented different ontological accounts that cast doubt on our intuitions by rejecting the existence of some objects that we usually assume to exist. Traditionally, the two most extreme positions are eliminativism, which states that no ordinary objects as compositions exist whatsoever, and universalism, which suggests that *any* objects jointly compose a new object.

In recent years, many scholars have begun to shift their interest toward metametaphysics, thus pursuing the overarching question of how metaphysics should be conducted in general. A new emphasis on the sciences has thereby evolved, most strongly shaped by the work of Ladyman and Ross. These two scholars have critically examined the contemporary state of metaphysics[1] and believe that "contemporary analytic metaphysics … fails to qualify as part of the enlightened pursuit of objective truth, and should be discontinued."[2] Arguably, the metaphysical enterprise has become detached from science in such a way that central arguments can no longer be well accounted for, given our contemporary understanding of physical nature. Therefore, Ladyman and Ross have suggested a new orientation toward a *scientific metaphysics* that embraces scientific findings, while emphasizing fundamental physics. Nevertheless, since the release of their book "Every Thing Must Go", the arguments for and against the existence of ordinary objects

---

1 Ladyman and Ross, *Every Thing Must Go*, §1.
2 Ibid., vii.

*Corresponding author: Arthur C. Schwaninger,** University of Zurich, Zurich, Switzerland; E-mail: arthur.schwaninger@uzh.ch

continue to persist in the traditional manner, as we have encountered them before this critique. This is exemplified by two recently published books by Korman and Zemack[3] and Benovsky[4] on ordinary objects, which provide a detailed overview of the contemporary debates, while reaching very different conclusions regarding these objects' ontological status. One possible reason that no overarching methodological shift has occurred is simply because being proficient in both the sciences and philosophy is difficult, and professional philosophers would have to readjust. However, as discussed in more detail below, a more important problem is that Ladyman and Ross have not provided a completely satisfying answer regarding how to approach the typical puzzles with which analytic metaphysicians have been concerned. Thus, the question asked in this paper is: how could a turn toward *scientific metaphysics* appear in terms of debates about ordinary objects?

I will argue that most recent developments in machine learning and computational neuroscience provide a new way to investigate the traditional problems we face in metaphysical disputes by analyzing and using simulations from the domain of computational neuroscience. In Section 2, I provide a brief overview of the ways analytic metaphysics has engaged with ordinary objects from within the manifest image, and which puzzles this approach generates. Section 3 then discusses the naturalist program in metaphysics, as presented by Ladyman and Ross. By more closely examining our contemporary understanding of fundamental physics, Section 4 discusses a naturalist approach to understanding ordinary objects and emphasizes the importance of building connections to our cognitive system. Section 5 introduces some contemporary views on cognition and neural processing, and discusses how computational models and simulations can provide answers to fundamental questions about the manifest image of the world. This paper seeks to help us progress in developing a unified picture of the world, given the best contemporary scientific theories about physics and the neurosciences.

2 The manifest and the scientific image: world views in conflict

Sellars argues that there is a discrepancy between our common sense understanding of the world and what science tells us about how the world *actually* is.[5] We think that the world consists of a vast number of living organisms and ordinary objects, such as chairs, tables, mountains, stars, and so forth. This is what Sellars calls the "manifest image" of the world. It corresponds to our "folk" picture of causation and mereological composition. In contrast, science presents an alternative picture of what exists in the world. In 1911, Ernest Rutherford's famous elastic scattering experiment with charged particles demonstrated that the material universe must be largely viewed as an empty void. Contemporary physicists present even stranger models that do not describe the universe as a composition of hunks of matter, but rather in terms of quantum fields that are mysteriously entangled.

Analytic metaphysics traditionally placed greater emphasis on the manifest than on the scientific image. One of the reasons for this asymmetry can be traced back to its beginnings with Quine. On the basis of Darwin's theory of evolution, Quine argues in favor of a naturalization of epistemology.[6] He states that the physiological configuration of the human organism must be equipped to access truth in a reliable way to sustain itself. A cognitive system that fools us about reality would have been replaced by natural selection, as it would not have been beneficial for survival. This argument sounds convincing and has led many subsequent philosophers to pursue metaphysical questions from a purely conceptual perspective, without relying on empirical evidence that extends beyond personal experience. This trend has also been influenced by the development of ordinary language philosophy, where Austin's[7] discussion of "the Nature of Reality" has led many philosophers to adhere to the position of naïve realism, according to which we have immediate access to reality.

3 Korman and Zemack, *Objects*.

4 Benovsky, *Eliminativism, objects, and persons*.

5 Sellars, *Science, Perception, and Reality*, §1.

6 Quine, *Ontological Relativity*.

7 Austin, *Sense and Sensibilia*.

Peter Strawson presents further influential arguments about why we must embrace the manifest image. Strawson's[8] "descriptive metaphysics," as a well thought-out ontological framework of what he claims to be most compatible with our daily intuitions, presents arguments against any non-object-based ontology that does not align with most of our daily experiences. He begins by observing that the way we think about the world involves assuming that there are separate, individual things that we can discuss and refer to. In the spirit of Kant's transcendental philosophy, Strawson investigates the underlying necessary conditions of this mode of thinking. Given that our thinking about the world is based on objects in a single spatiotemporal system and we are able to both identify and re-identify these, he argues that the actual existence of these objects in the world is required for us to perceive the world the way we do. Further, he argues that interpersonal communication about objects would fail if these objects do not actually exist.

Although Strawson's argument has been criticized as unsound,[9] it has become very influential and is commonly accepted as a hurdle for any non-object-ontologist to overcome. One of the few proponents of process philosophy, Nicholas Rescher, points out that identification is a (cognitive) process itself, and argues that (re-) identification is ultimately an arbitrary, unqualified criterion for ontology.[10] Nonetheless, Strawson's argument ultimately follows from linguistic considerations that words are the basis of all logical reasoning. For him, "identifying reference" is essential to human thinking and speech.

In late twentieth-century analytic philosophy, many philosophers found numerous puzzles and paradoxes within the manifest image that seem to cast doubt on the consistency of these conservative views. Believing that chairs and houses exist raises the following commonly cited puzzles:

(a) The sorites paradox: (P1) One stone is not a heap. (P2) If a pile of N stones does not constitute a heap, then neither does a pile of $N + 1$ stones. (C) Therefore, one million stones do not constitute a heap.

(b) The problem of material constitution: (P1) A statue and the lump of clay from which it is formed have different properties. (P2) If so, then, by Leibnitz's law,[11] the statue and lump are not identical, and constitute distinct coincident objects. (P3) Distinct coincident objects cannot exist. (C) Therefore, the statue does not exist.

(c) The problem of the many: How can I say that I see *one* cloud in the sky, when in fact a multiplicity of subsets of droplets would each constitute a different cloud?

One possible way of responding to these types of puzzles is offered by eliminativism–a view that rejects the concept of composite objects and argues that ordinary objects should be eliminated from our ontology.[12] Eliminativism gained great attention a decade ago, and still attains some degree of popularity. By endorsing the principle of parsimony and avoiding any causally redundant entities, eliminativism has been argued by its proponents to best align with our scientific understanding of the world.[13] Merricks, who strongly advocates this view, argues that his position is the most compatible with scientific views, regardless of what physical theories state. However, as discussed later, some philosophers have urged that this is not the case and that Merricks's assumptions about physics are incompatible with contemporary fundamental physics. Nonetheless, his views remain instructive to understand how we can conceptualize reality only from within the manifest image, given the values of science. Merricks explains his position most intuitively in the following manner.[14] Let us assume there are N atoms in a room arranged *statue-wise*. In that case, there are N objects in the room. However, if we assume that statues actually exist, there would not be N, but rather $N + 1$ objects in the room, where one of these objects is *identical* to N others by reduction. Merricks

8 Strawson, *Individuals.*

9 Stroud, "Transcendental Arguments."

10 Rescher, *Process Philosophy,* 40.

11 For extensional objects, Leibnitz's law of indiscernibility states that two objects are identical if and only if they have all their properties in common.

12 For a detailed discussion, see Korman and Zemack, *Objects.*

13 Merricks, *Objects and Persons.*

14 Ibid., §1.

states that this is not a plausible position because a one-to-many identity relationship[15] cannot hold, as it entails mereological essentialism, which is typically rejected by straightforward analysis. In other words, to assume the existence of ordinary objects means to assume composition as identity. Given that the latter assumption is false, there are no ordinary objects from this scientific perspective.

Eliminativism does, however, remain an unpopular view these days. Many philosophers deny that judgment about ordinary objects is an actual part of scientific discourse,[16] and follow the views of Stebbing, who says: "I venture to suggest that it is as absurd to say that there is a scientific table as to say that there is a familiar electron or a familiar quantum."[17] However, these philosophers are inclined to state that ordinary objects might not exist in a scientific manner, but in some other sense. Even if they do not exist at a fundamental physical level, they can still be said to exist in the ordinary sense of language usage, outside the ontological seminar room.[18] However, other philosophers view the aforementioned puzzles as more pressing. They view them as deep philosophical challenges to widely used metaphysical concepts, and consider philosophers who believe that these puzzles generate a "merely" linguistic problem are missing the point. For example, Heller argues that the "[s]orites paradox shows us that the world is not as we think it to be and challenges us to give an account of just what the world is like."[19]

The debates above place the manifest image as their starting point of philosophical reflection. For example, Merricks's argument rests on the assumption that an ideal and complete physics will postulate individuals that are the simples from which all other material objects to be composed. However, some more recent philosophers have insisted that these underlying assumptions are inconsistent with contemporary fundamental physics. This topic is further discussed in the next two sections.

3 Scientific metaphysics and the Principle of Naturalistic Closure

 In more recent times, we find a new trend in metaphysics that no longer focuses on linguistic or conceptual analysis, but rather approaches metaphysical questions from a science-centered perspective. This new movement of radically naturalizing metaphysics was initially most strongly promoted by Ladyman and Ross,[20] and is commonly referred to as "scientific metaphysics." What characterizes this trend is its refusal to appeal to intuition to motivate a priori claims, while simultaneously—in contrast to the (logical) empiricists, such as Hume or the Vienna Circle—preserving a positive metaphysical view. Although, as previously discussed, Merricks argues that his views conform to *any* fundamental physical theory, Ladyman and Ross[21] do not take such arguments seriously. They ridicule any micro-reductionist arguments because these arguments make presuppositions about reality that do not have any grounding in contemporary quantum theory, yet rather correspond to an understanding of physics "learned in A-level chemistry."[22] In disagreement with Oppenheim and Putnam's[23] foundationalism about physics, Ladyman and Ross also propose that scientific metaphysics ought to focus on bringing forward a unified picture of the independent scientific disciplines by following two main principles. First, we should follow what they call the *Primacy of Physics Constraint* (PPC), stating that:

> Special science hypotheses that conflict with fundamental physics, or such consensus as there is in fundamental physics, should be rejected for that reason alone. Fundamental physical hypotheses are not symmetrically hostage to the conclusions of the special sciences.[24]

---

15 Merricks follows the relational definitions of Lewis, *Parts of Classes*.
16 Thomasson, "Ontological Minimalism."
17 Stebbing, *Revival: Philosophy and the Physicists*, 58.
18 van Inwagen, *Existence*.
19 Heller, *The Ontology of Physical Objects*, 69.
20 Ladyman and Ross, *Every Thing Must Go*.
21 Ibid., §1.
22 Ibid., 24.
23 Oppenheim and Putnam, "Unity of Science as a Working Hypothesis."
24 Ladyman and Ross, *Every Thing Must Go*, 44.

This ensures that the naturalist respects the empirical sciences and is in discord with dualist and emergentist positions. The second principle they endorse is the *Principle of Naturalistic Closure* (PNC):

> Any new metaphysical claim that is to be taken seriously at time *t* should be motivated by, and only by, the service it would perform, if true, in showing how two or more specific scientific hypotheses, at least one of which is drawn from fundamental physics, jointly explain more than the sum of what is explained by the two hypotheses taken separately.[25]

From the perspective of unifying the sciences, both principles are motivated by the fact that physics has the widest scope compared to all the other sciences. To achieve a sort of unification, the maximum scope must persist in any greater theory. The PNC allows for and endorses metaphysical claims that build bridges between "separately developed and justified pieces of science (at a given time) [that] can be fitted together to compose a unified world-view."[26]

4 Making sense of ordinary objects

To achieve the envisioned scientific unification for which Ladyman and Ross hope, we must also make sense of ordinary objects, as these are a central part of our empirical investigations and thoughts. Peter Strawson writes: "We think of the world as containing particular things some of which are independent of ourselves."[27] Even if we reject Strawson's metaphysics, our phenomenological experiences of *things* must be accounted for in a naturalist metaphysics, at least because many of the special sciences discuss and formulate theories about classes of ordinary objects. Failing to account for the daily experiences we make of a world consisting of chairs, tables, and trees simply cannot be said to provide a unified picture of the world[28].

Daniel Dennett emphasizes this point when he writes about scientific metaphysics. He says that "at least a large part of philosophy's task … consists in negotiating the traffic back and forth between the manifest and scientific images."[29] In this sense, philosophy since antiquity has posed challenges to a unified understanding by identifying puzzles such as those mentioned earlier. From the naturalist's perspective, it is our task today to account for these puzzles in one unified theory.

This naturalist approach of unifying the science into one greater picture and reconciling it with the manifest image is not uncontroversial. Rosenberg argues that throughout history any attempts of unification have failed, which is a good indicator for this project being futile.[30] To understand some of the difficulties we face, let us first examine the status of ordinary objects in contemporary physics.

Even though classical Newtonian mechanics was once found to be highly unintuitive, most of all due to Newton's concept of force, it is nonetheless a theory that can explain and makes predictions that are largely in accordance with our ordinary, everyday experiences. Newton's theory of matter essentially conforms to Democritus's atomism which remains the basic intuition about matter for many non-physicists today. If physics had ended here, we may not have to worry about providing an additional account of ordinary objects beyond the perspectives of Merricks and others. However, the issue has been complicated by quantum mechanics (QM), at least if we suppose that the features peculiar to QM not only concern the microphysical realm, but also extend to the macroscopic world. Under this supposition, we encounter highly perplexing problems when seeking to apply the notions of QM to macroscopic material objects. For example, this is illustrated by the *measurement problem* most prominently discussed in terms of Schrödinger's cat.[31] I do not intend to address all these different problems, but only focus on some issues discussed in the context

25 Ibid., 37.

26 Ibid., 45.

27 Strawson, *Individuals*, 15.

28 Hofweber, "Empirical Evidence and the Metaphysics of Ordinary Objects," even goes as far as to say that we have empirical evidence for the existence of ordinary objects and some realms of science could not operate without.

29 Dennett, "Kinds of Things", 99.

30 Rosenberg, *The Atheist's Guide to Reality*, 218.

31 Schrödinger, "Die gegenwärtige Situation in der Quantenmechanik."

of quantum entanglement—namely, the issues of localization, separation, and identification of ordinary objects.

The entanglement of particles describes the correlation of their fundamental properties that does not occur by chance. If one adopts a realistic approach to QM, then QM and classical mechanics cannot be aligned by the inclusion of hidden variables, and we are committed to regarding quantum entanglement as a feature of nature that exists objectively, independent of the observer. On this basis, without going into detail, quantum entanglement conflicts with three central assumptions we usually make about ordinary objects.[32] First, contrary to our intuitive classical view of the world, quantum systems and their properties are generically not *localized* at a specific region in space-time. Second, quantum systems cannot be described in isolation from other systems when entangled. On quantum mechanical grounds, this issue defeats the classical *spatial separability principle*[33] by which the state of each individual system determines its local properties. As a result of quantum entanglement, the system's local properties are thought to be determined by the joint state of the entangled particles. Third, entangled quantum particles are not distinct *individuals,* as there exists no property that allows us to distinguish one from another.[34] If one assumes that the properties of entangled systems in QM are correctly interpreted to not supervene on their parts and that quantum entanglement is not limited to the microscopic realm, but propagates to macroscopic systems, then some monists have reached the conclusion that the whole is prior to its parts.[35] Under the assumption that everything interacted during the Big Bang, there is good reason to believe that all particles are entangled and build one universal system.[36] It remains unclear whether everything is entangled or not; however, it is implausible to assume that the ordinary objects of our experience can ontologically be clearly distinguished from each other on the basis of physics.

If we believe that QM is approximately true and that the principles of localization, separability, and individualization have no physical basis, then any other theory that does postulate ordinary objects in the classical sense will conflict with our most fundamental scientific understanding. However, based on the PPC, how can a unified picture of the world be achieved, if ordinary objects are so vital to our understanding of the world? The answer provided by Ladyman and Ross is to treat ordinary objects not as individuals, but as "real patterns"—a term borrowed from Dennett[37] that refers to relatively stable and enduring structural patterns within the data that allow for efficient description.[38] What we perceive as individuals are, for Ladyman and Ross, "only epistemological book-keeping devices"[39] that are constructs of our cognitive system. Thus, instead of postulating the existence of *individual* ordinary objects, we can explain how our experiences of ordinary objects as individuals *come about* in a way that does not conflict with science. As stated by Dennett:

> The ontology of everyday life is now teeming with items that, like fatigues, sit rather awkwardly in the world of atoms and molecules. If we can understand how this population explosion *came about*, and why it is so valuable to us as agents in the world, we can perhaps discharge our philosophical obligations without ever answering the ultimate ontological question.[40]

For Ladyman and Ross, different real patterns are found and exist on different scales. The table I am working on is a real pattern that exists at the scale of ordinary human perception—even if the table is not actually an individual, but is only cognitively perceived as such. However, Ladyman and Ross claim that it

---

32 These three points are elaborated in much greater detail in Esfeld, "Holism in Cartesianism and in Today's Philosophy of Physics"; "Quantum Holism and the Philosophy of Mind."

33 Howard, "Holism, Separability, and the Metaphysical Implications of the Bell Experiments."

34 French and Redhead, "Quantum Physics and the Identity of Indiscernibles."

35 Maudlin, "Part and Whole in Quantum Mechanics"; Esfeld, "Quantum Holism and the Philosophy of Mind."

36 Gribbin, *In Search of Schrödinger's Cat*, 229.

37 Dennett, "Real Patterns."

38 The representation of any data is assumed to be *efficient*, as long as its description requires less storage space than the entire bit map of the raw data.

39 Ladyman & Ross, *Every Thing Must Go*, 240.

40 Dennett, "Kinds of Things", 106; emphasis added.

makes no sense to say that the table exists at cosmic or quantum scales. This ensures that the ontological commitments of the special sciences are on a par with those of physics, and there is no ontological superiority of physics over the other sciences. The only difference between the two scientific domains is that the *laws* of fundamental physics constrain the special sciences by the PPC, and not the other way around. This results in an asymmetry among the sciences regarding the priority of the laws they postulate, while also maintaining all claims concerning existence on equal grounds. Thus, while the *laws* of fundamental physics enjoy a status of priority, the *entities* postulated are no "more real" than tables and chairs. The patterns of tables and chairs are "real" in the sense that they are not merely subjective or mind dependent, and their description cannot be further compressed information-theoretically.

Harman critically examines the ontology proposed by Ladyman and Ross and questions how a multiplicity of patterns and scales comes about.[41] He says that the division into discrete scales and the existence of a plurality of real patterns is inconsistent with an objective, autonomous reality. I agree with Harman that further explanations of the discretization of reality are required, yet I deny that this is impossible. In their unifying metaphysics, Ladyman and Ross should include a theory that allows complex systems, such as human beings, to think of the world as a composition of individuals. To understand why it is necessary for a human observer to discretize reality the way we do, I encourage naturalist philosophers to examine how our perceptive system operates and what makes us perceive reality as a collection of individual things that we can track over time.

The way Ladyman and Ross discuss real patterns and the way they understand how these are perceived in the mind as individuals, can be traced back to Dennett's original paper in which he writes:

> [A pattern is] discernible to the naked human eye ... because of the particular pattern-recognition machinery hard-wired in our visual systems—edge detectors, luminance detectors, and the like.[42]

This quotation expresses how we thought about perception and our visual system a decade ago, when a representational theory of the brain still dominated the neurosciences. Today, however, the doctrine has shifted toward a Bayesian view of the visual system—often in connection to what is called the "predictive processing theory of the brain."[43] Based on this view, it is not the case that the brain merely detects and recognizes distinct patterns simply because of some hard-wired feature-detection mechanisms, and, with this, we do not see a discrete entity without possessing a prior, internal model of the world. Only a complex system that has already gained knowledge about the world via a dynamic learning process can recognize real patterns as a plurality and distinguish them from each other. To unify physics with the special sciences through the PNC, we must take these learning processes into account.

What I present here is by no means the first approach to ordinary objects from a machine learning or neuroscience point of view. For instance, Brian Smith offers a metaphysics that retains our common sense realist intuition of a mind independent world, while preserving a constructivist stance that suggests that the specific computations of the mind single out one possible symbolic representation from a multiplicity of alternatives. The account lays out the foundations of a metaphysical theory of objects, by which ontology, representation, and intentionality are intrinsically interconnected. Any statement about the world depends on representation, and any attempt of representation brings the world's objects into existence together with their properties. Thus, talking about the world is an act of object-making, says Smith. The theory Smith offers can be seen as important foundational work for a metaphysical approach called "Cognitive Metaphysics" that was recently encouraged by Decock[44] and will be further discussed below.

Given that there have been some major developments in machine learning and the neurosciences over the past decade, I belief it is necessary for metaphysicians to engage with these most current findings.

---

41 Harman, "I Am Also of the Opinion That Materialism Must Be Destroyed."
42 Dennett, "Real Patterns", 33.
43 See, e.g., Hohwy, *The Predictive Mind*; Wiese and Metzinger, "Vanilla PP for Philosophers"; Clark, "Whatever Next? Predictive Brains, Situated Agents, and the Future of Cognitive Science."
44 See Decock, "Cognitive Metaphysics." The envisioned new realm of metaphysics goes back to Brown, "Foundations of Cognitive Metaphysics."

Ideally, this approach will allow us to make statements about our direct experiences that are consistent with the physicist's view of the world. The concepts of separability and identification are crucial to our ordinary common sense understanding of the world; thus, explaining how the many objects we experience in the world arise first requires an explanation of how we experience independent, isolated objects that are separated from each other and can be identified as such, without initially assuming their ontological independence. In particular, to do justice to Strawson's work, we are expected to explain the phenomenon of (re-) identification and communication without presupposing the concept of ordinary objects in the first place. This is the challenge we face, and, in the following, I argue for an approach to the challenge that considers the results from simulations performed in contemporary machine learning and computational neuroscience.

5 Moving toward a computational understanding of the manifest image

During the last few years, our knowledge in machine learning and information processing has expanded considerably. The introduction of Deep Learning techniques,[45] such as Convolutional Neural Networks, has allowed us to develop methods and algorithms for a wide variety of different tasks, including object classification, representation, and segmentation, and image generation. These new technological accomplishments are the results of creating complex neural network architectures where each individual artificial neuron has very simple properties, yet the network as a whole can *learn* from the sensory data. When such artificial neurons are presented with some sensory information, the current hypothesis of the leading experts in the field is that the neurons collectively perform *inference* by changing their state to better "explain" the observed data.[46]

Theoretical neuroscience and machine learning research have always been closely related. Recently, there have been numerous publications suggesting that deep learning methods—which have been developed for engineering purposes—operate similarly to the way the brain processes information.[47] In combination, both machine learning and the neurosciences are generating answers to the fundamental questions of how humans are able to create and identify objects, in a way that they can be classified, placed into relationships, and semantically processed.

Given that neural networks can learn their own categories and identify objects in images, I argue that a naturalist account through the eyes of machine learning and theoretical neuroscience simulations can provide answers to *seemingly* purely metaphysical question about ordinary objects. If we assume that QM offers little reason to believe that the individual objects of the manifest image exist in the fundamental sense, then the most reasonable explanation for our experience of such separable objects is that a certain cognitive process must construct them from the flow of registered information. As stated by Decock, we can either interpret the puzzles about ordinary objects as purely epistemic questions in metaphysical disguise, or we can believe that epistemic and metaphysical questions are entangled in a Kantian framework.[48] Exploring the computations performed by the brain can help us make sense of certain features of the manifest image—particularly how the cognitive system shapes our beliefs concerning tables and chairs. To provide evidence for these views, I will demonstrate how simulations answer the following three questions regarding ordinary objects:

---

45 Goodfellow, Bengio, and Courville, *Deep Learning*.
46 Hinton and Sejnowski, "Learning and Relearning in Boltzmann Machines"; Friston and Stephan, "Free-Energy and the Brain"; Berkes and Orbán, "Spontaneous Cortical Activity Reveals Hallmarks of an Optimal Internal Model of the Environment".
47 Sacramento et al., "Dendritic Cortical Microcircuits Approximate the Backpropagation Algorithm"; Sacramento et al., "Dendritic Error Backpropagation in Deep Cortical Microcircuits"; Bengio et al., "Feedforward Initialization for Fast Inference of Deep Generative Networks Is Biologically Plausible"; Scellier and Bengio, "Equilibrium Propagation: Bridging the Gap between Energy-Based Models and Backpropagation."
48 See also Swanson, "The Predictive Processing Paradigm Has Roots in Kant," who argues that predictive processing echoes Kant's approach to metaphysical questions.

1. Why are ordinary objects perceived as separate, individual things, if they do not actually exist in this physical sense? (Strawson)
2. If objects do not exist in the sense of being separable, individual entities, how can two independent agents communicate with each other about them? (Strawson)
3. If all things are somewhat mental, how can real things be different from imagined ones? How can we distinguish between fictitious and real objects?[49]

5.1 Explaining the experience of separability

In the following, let us examine simulations that can help explain how the perception of separate objects is possible. Most recent developments in computational neuroscience do not conceive the brain as passively computing sensory information with predetermined feature-detection mechanisms. Instead, the brain is viewed as actively engaged in forming the best hypotheses it can infer about the environment (given some computational constraints, as explored below). This is computationally realized by having a "recognition model" of the environmental that processes the upstream information flow from the sensory input to the higher cortical regions of the brain. Meanwhile, the recognition model is constantly updated by a "generative model", that generates images by processing information from higher to lower cortical regions of the brain. These generated images are the predictions of higher cortical regions about lower cortical activity. Any mismatch in the prediction updates the recognition model. Figure 1 displays three layers of a neural network, with each layer having information flowing in both directions. In this figure, I only indicate what a single neuron cell "sees," which is the information flowing both upwards and downwards.
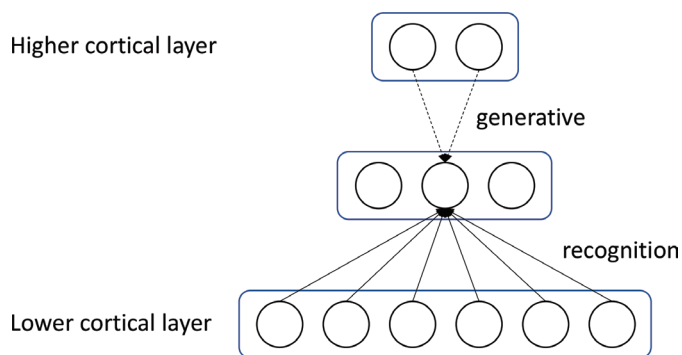

Figure 1: Each hidden neuron of the network receives information from higher and lower cortical layers (here only indicated by the central neuron). The model running from top to bottom is a generative network that produces images, whereas the model from bottom to top is the recognition network.

This concept can best be illustrated by a toy example shown in Figure 2. This example has data (2a) that consist of images that humans perceive as either "0" or "2." We perceive these images as such because our neurological system has already created an internal representation of these categories. However, each of the data's instances only consists of a number of pixels with different pixel values, and there is no clear separation between the zeros and twos. Nonetheless, some neural network models (2b) can make clear separations of the data over the course of a learning process. This can be achieved in a semi-supervised manner, without any labeling of the input images. In the case discussed here, the input image is also set to be the output target image (in an auto-encoder fashion). Therefore, the objective of the network is to recreate the input values as successfully as possible by building an internal representation of the data (2c) and generating an image from this representation. The "goodness" is determined by an objective function—a function that returns a high value when the model's input and generated output strongly diverge. This function should be minimized by learning an adequate representation of the data.[50]

---

49 This is a common response we already find against 18th century idealism. See e.g. McDonough, "Berkeley on Ordinary Objects" and Downing, "George Berkeley".
50 This is similar to how we think the brain learns about the world. See details in Rao and Ballard, "Predictive Coding in the Visual Cortex."

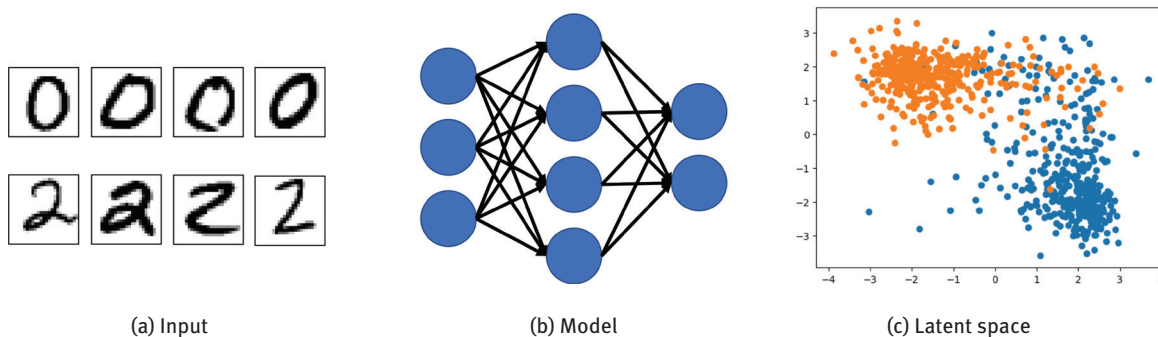(a) Input   (b) Model   (c) Latent space

Figure 2: The raw input data (a) are transformed by the model (b) into two separate clusters in the network's representation or latent space (c). Those points that are far away from the cluster centers correspond to images that have been presented to the model at an early stage in the learning process, where the model has not yet managed to build its internal categories.

Based on such learning techniques, unsupervised image segmentation algorithms have become very popular in the computational vision community.[51] After learning an internal representation of the world, these algorithms can distinguish an object from its background. Given such a representation, the usual approach is to compute for each pixel the probability of being connected to the surrounding pixels. The upper section of Figure 3 displays some colored images of objects that individuals can encounter in their daily lives. Below, these objects are distinguished from their backgrounds algorithmically. Within the machine learning community, there exist a huge number of different models of how such an algorithm can be implemented in supervised or unsupervised fashions. It is unknown which of these methods most closely corresponds to the algorithm that our brains use when performing image segmentation. Perhaps the techniques that our brains employ are much more sophisticated than any of the existing models, or perhaps it is only the brain's vast computational resources that allow it to perform object segmentation with such ease.
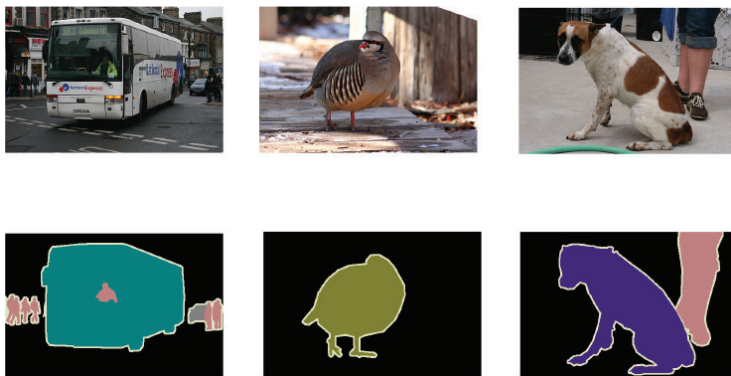

Figure 3: A semantic image segmentation algorithm uses the upper images as input and returns the lower images as output.[52]

Whatever the underlying computational mechanisms of the brain are, we learn from these simulations that we need not assume the existence of objects to make sense of image segmentation. A system that perceives the world by segmenting images will be inclined to believe that there fundamentally exist separable ordinary objects or substances, even if this is actually not the case.

Returning to Quine, it seems initially puzzling that the existence of separate, self-subsistent objects do not exist in the physical world. As stated by Quine, evolution must have enabled us to perceive the world as it is to maximize our chances of survival. Therefore, we would assume that ordinary objects have a physical

---

51 For a survey on unsupervised segmentation algorithms, see Zhang, Fritts, and Goldman, "Image Segmentation Evaluation." See also Spratling, "Image Segmentation Using a Sparse Coding Model of Cortical Area V1" for an image segmentation model of the cortical area V1.

52 Images from Chen et al., "Semantic Image Segmentation with Deep Convolutional Nets and Fully Connected CRFs."

basis independent of our minds; however, this conclusion is drawn too quickly. Simulations conducted in computational neuroscience research provide strong reasons why there is an advantage for organisms to see objects as individuals. It is not possible to go into technical detail here, but I will briefly discuss some results that arguably indicate how an evolutionary benefit derives from experiencing the world as we do.

By sampling the world through the sensory system, the brain maps a continuous probability distribution onto a lower dimension space. I illustrate this implicitly in Figure 2, where each image (2a) has the size of 28 x 28 = 784 pixels, and is thus a point in a 784-dimensional space. The data consisting of a set of images of zeros and twos are samples from probability distributions in this very high-dimensional space. Computationally speaking, the brain has two options: it either explicitly encodes for an approximation of this probability distribution or it encodes for the distribution's different states. Today, it is strongly believed that the brain does not encode for the full probability distribution, but only for the different states expressed by the distribution's peaks. This is computationally realized by introducing a non-linearity (e.g., threshold) that distinguishes the different peaks of the probability distribution (e.g., based on the local maximum likelihood).

By simulating point neurons, Petrovici[53] computationally explains our experiences when we are confronted with certain optical illusions. Consider the famous case of the duck-rabbit. A subject looks at the duck-rabbit image and alternates between two different image interpretations. This situation is considered to indicate that the network activity encodes for states, rather than probabilities. The brain is either in one state or another, and is not in an intermediate state. If the reverse were the case, then we would not see either a duck or rabbit—rather, our brain would judge the duck-rabbit to be a superposition of both a duck and a rabbit, with different probabilities (e.g., 40% duck and 60% rabbit). There are at least two simple reasons why such a neural implementation would not be useful for survival. First, as Petrovici demonstrates, computing multiple possible states is computationally very inefficient,[54] resulting in wasted resources. Second, the brain must be committed to a decision to act in response to one of the categorical alternatives indicated by the peaks in the distribution of the probability mass function. If the visual cortex computes that there is a 10% chance of a donkey, 30% chance of a house cat, and 60% chance of a hungry lion standing next to me, then there is little benefit in adjusting my behavior according to the first two probabilities. Therefore, bringing the brain into a determinate state was favored over the course of evolution.

5.2 Explaining communication

Once a model of an individual perceiving unit—or an observing agent—is trained, simulations of agents communicating about objects are possible. This is achieved in the following manner. After training each agent's model, the individual network can identify an object and generate an own image from the corresponding class. This self-generated image is shared with the other agent, who ideally identifies the image correctly. More precisely, the simulation setup is as follows (see Figure 4). The model consists of two agents, A and B, who are represented by two slightly different neural network architectures. Both agents have been trained on different sets of the image database to illustrate that communication does not require the exact same prior experience. When A chooses a sample point within a region of high density in its latent space, it generates an image via top-down computations over the network. This image is shown to B, who *identifies* the image by forward, upstream propagation, and samples a new self-generated image from the corresponding class by top-down generative image processing. This alternation between generation and identification of images is a process that allows agent A to be informed that agent B has sampled the correct internal class representation, and vice-versa. Thus, both agents possess the information that they are both communicating about the same object, even if no such *thing* actually exists, independently of their joint construction.

---

53 Petrovici, "Form Versus Function", §6.
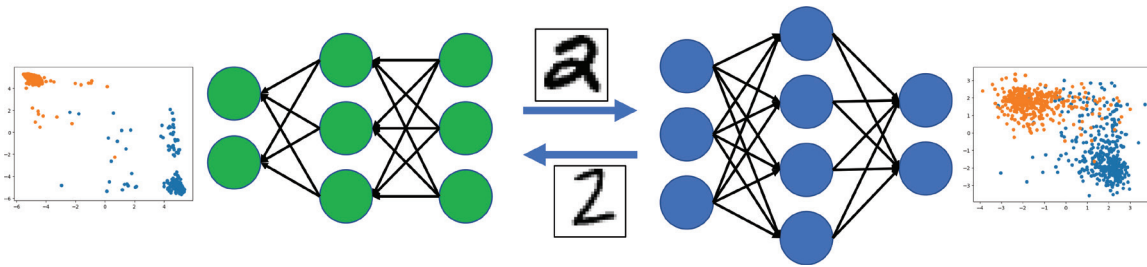**54** Ibid., §6.4.

Figure 4: Two trained neural networks can communicate with each other by receiving generated images from the other network and returning images of the same class.

Simulations such as those presented here indicate that typical arguments against non-object ontologies fail. There need not be any-*thing* out there that is a precondition for my capacity to communicate and experience the world as I do, in terms of separate individual objects. The only crucial points are that *all communicating agents share a similar cognitive process* with regard to the underlying computations and are exposed to a similar environment. This enables the agents to understand each other, if communication is modeled as described above. A direct consequence is that this defeats our natural inclination to a universal object ontology. For most philosophers, it "seems plausible to suppose that there is nothing we can say that the Martians can't,"[55] and vice-versa. However, this view is incorrect, based on the perspective provided here. If the Martians' information-processing system is encoded very differently, it will not necessarily be able to have the same categories as humans do, and will not necessarily individuate the same objects. In principle, learning theory tells us that it is possible for a Martian to never grasp the concept of a teacup, regardless of how many cups I present it with, and that the Martian will not be able to recognize the cup as an individual, given only its perceptual system.[56] This is in line with Smith, who suggests that individuation is dependent on an organism's cognitive processes.[57] Nonetheless, I should note that if the Martian is a technically well-equipped scientist with an understanding of mathematics and statistics, the human could still communicate the idea of a cup even if the Martian cannot train her perceptual system to "see" it as an individual by looking at it.

5.3 Distinguishing fictitious from real objects

The semi-supervised model used above to explain communication involves both a generative and recognition component. The recognition model is capable of identifying and re-identifying the objects it perceives. Section 5.1 showed that the generative network model creates images that seek to predict the neural activity of lower cortical regions as closely as possible. The purpose of this is to improve the recognition model. More precisely, the generated images can be used to learn the recognition model by relying on an objective function that minimizes the error between the input and self-generated image. Interestingly, however, the generative network can also be used for other purposes, such as "dreaming" new, never-before-perceived images. This is achieved by (partially) turning off the recognition network and letting the generative model dominate the network. As a consequence, the generative component can create new *dreamed* instances of an existing category to make predictions on the sensory data.

   Generative models are the key to understanding dreaming and the imagination of fictional entities. To provide an example, Figure 5 displays some images that were *imagined* by generative neural networks. Looking at the faces in Figure 5b, we find that the human eye cannot distinguish these imagined faces from real ones. Although the fictitious images are not distinguishable from real-world sensory input, the neural network behind these images can still easily distinguish between top-down and bottom-up information

---

55 van Inwagen, "Being, Existence, and Ontological Commitment", 479.
56 This point is ultimately guaranteed by the No-Free-Lunch-Theorem stating that no universal learning algorithm exists. See Wolpert and Macready, "No Free Lunch Theorems for Optimization."
57 Smith, *On the Origin of Objects*; Smith, "Reply to Dennett."

flow, either by rule-based or learning-based operations. For example, this is achieved in in Generative Adversarial Networks, where one network generates images, trying to fool another network that is trained to discriminate self-generated images from real images.[58]

(a)   (b)

Figure 5: Different images created by a generative neural network. (a) An image of Descartes in the style of Van Gogh's paintings, based on the model of Gatys et al. (2016). (b) Faces of non-existent people.[59]

The predictive processing theory of the brain suggests that information is not only processed upstream from lower-level cortical regions that are closer to the sensory input and moved to higher cortical regions, but also vice-versa. The theory states that this process does not simply occur occasionally, but occurs continuously. In the case in which discriminating, bottom-up information flow dominates the generative net component, the brain projects less prior structure onto lower cortical regions, thereby inflicting only small changes on the data. In the reverse case, however, the input data might get completely distorted by the generative network. The brain can generally control the degree to which things are imagined. Under normal circumstances—where the subject is neither dreaming, taking psychedelic drugs, nor suffering from certain mental disorders—the imagined and the real are neurologically distinguished by the brain's attention to information flow. The significance of simulations here is that they provide proof for the possibility of such a network, and subsequently help answer the question of how fictional objects are distinct from "real" objects.

6 Conclusion

Through using simulations, we can offer explanations about how neural processes build an internal representational model of the acquired sensory data. By investigating these simulations more closely, we can determine how these computations explain some of our experienced features of the world. Over the course of the simulation, the neural model creates clusters from which we can explain how distinct categories and separate objects arise in the visual cortex, given the data that the model was fed. As the brain seeks to maximize its internal order, while computing for states, rather than probabilities, the subject must experience the world to consist of ordinary objects, such as chairs and tables. This is the world that causes philosophers to ponder over puzzles in the manifest image regarding perception, epistemology, and metaphysics. The process that creates these experiences can be expressed in a set of mathematical operations or a sequence of instructions computable by the brain. These equations or instructions only provide a limited idea of how and why the world, as we experience it, comes about in a certain way and not another. Simulations allow us to obtain a new understanding of which properties emerge in the manifest image. They help us explore the manifest image in a different way to the philosopher's traditional approach of evaluating one's experiences as linguistic terms. Simulations help us visualize the categories that the network creates, and thus indicate to some extent how a person with the same visual input might think about the world.[60]

---

58 See details in Goodfellow, "NIPS 2016 Tutorial"; Goodfellow et al., "Generative Adversarial Nets"; Radford, Metz, and Chintala, "Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks.".

59 Images from Karras, Laine, and Aila, "A Style-Based Generator Architecture for Generative Adversarial Networks."

Bibliography

Austin, John Langshaw. *Sense and Sensibilia*. Oxford University Press, 1962.

Bengio, Yoshua, Benjamin Scellier, Olexa Bilaniuk, Joao Sacramento, and Walter Senn. "Feedforward Initialization for Fast Inference of Deep Generative Networks Is Biologically Plausible." *ArXiv Preprint ArXiv:1606.01651*, 2016.

Benovsky, Jiri. *Eliminativism, Objects, and Persons: The Virtues of Non-Existence*. 1 [edition]. Routledge Studies in Metaphysics 13. New York: Taylor & Francis, 2018.

Berkes, Pietro, and Gergő Orbán. "Spontaneous Cortical Activity Reveals Hallmarks of an Optimal Internal Model of the Environment." *Science* 331, no. 6013 (January 7, 2011): 83–88.

Brown, Jason W. "Foundations of Cognitive Metaphysics." *Process Studies* 27, no. 1 (1998): 79–92.

Chen, Liang-Chieh, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L. Yuille. "Semantic Image Segmentation with Deep Convolutional Nets and Fully Connected CRFs." *ArXiv:1412.7062 [Cs]*, December 22, 2014. http://arxiv.org/abs/1412.7062.

Clark, Andy. "Whatever next? Predictive Brains, Situated Agents, and the Future of Cognitive Science." *Behavioral and Brain Sciences* 36, no. 03 (June 2013): 181–204.

Decock, Lieven. "Cognitive Metaphysics." *Frontiers in Psychology* 9: 1700, (2018).

Dennett, Daniel C. "Kinds of Things—Towards a Bestiary of the Manifest Image." In *Scientific Metaphysics*, edited by Don Ross, James Ladyman, and Harold Kincaid. Oxford University Press, 2013.

Dennett, Daniel C. "Real Patterns." *The Journal of Philosophy* 88, no. 1 (1991): 27–51.

Downing, Lisa. "George Berkeley." In *Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, 17. Winter 2018 Edition, 2004. https://plato.stanford.edu/archives/win2018/entries/berkeley/.

Esfeld, Michael. "Holism in Cartesianism and in Today's Philosophy of Physics." *Journal for General Philosophy of Science* 30 (1999): 20.

Esfeld, Michael. "Quantum Holism and the Philosophy of Mind." *Journal of Consciousness Studies* 6, no. 1 (1999).

French, Steven, and Michael Redhead. "Quantum Physics and the Identity of Indiscernibles." *The British Journal for the Philosophy of Science* 39, no. 2 (1988): 233–46.

Friston, Karl J., and Klaas E. Stephan. "Free-Energy and the Brain." *Synthese* 159, no. 3 (November 12, 2007): 417–58.

Gatys, Leon A., Alexander S. Ecker, Matthias Bethge, Aaron Hertzmann, and Eli Shechtman. "Controlling Perceptual Factors in Neural Style Transfer." *ArXiv:1611.07865 [Cs]*, November 23, 2016.

Goodfellow, Ian. "NIPS 2016 Tutorial: Generative Adversarial Networks." *ArXiv Preprint ArXiv:1701.00160*, 2016.

Goodfellow, Ian, Yoshua Bengio, and Aaron Courville. *Deep Learning*. Adaptive Computation and Machine Learning. Cambridge, Massachusetts London, England: The MIT Press, 2016.

Goodfellow, Ian, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. "Generative Adversarial Nets." In *Advances in Neural Information Processing Systems*, 2672–2680, 2014.

Gribbin, John. *In Search of Schrödinger's Cat - Quantum Physics and Reality*. New York: Bantam Books, 1984.

Harman, Graham. "I Am Also of the Opinion That Materialism Must Be Destroyed." *Environment and Planning D: Society and Space* 28, no. 5 (October 2010): 772–90.

Heller, Mark. *The Ontology of Physical Objects: Four-Dimensional Hunks of Matter*. New York: Cambridge University Press, 1990.

Hinton, Geoffrey, and Terrence J Sejnowski. "Learning and Relearning in Boltzmann Machines." *Parallel Distributed Processing: Explorations in the Microstructure of Cognition* 1, no. 282–317 (1986): 2.

Hofweber, Thomas. "Empirical Evidence and the Metaphysics of Ordinary Objects." In *The Nature of Ordinary Objects*, edited by Javier Cumpa and Bill Brewer, 1st ed., 27–47. Cambridge University Press, 2019.

Hohwy, Jakob. *The Predictive Mind*. First edition. Oxford, United Kingdom ; New York, NY, United States of America: Oxford University Press, 2013.

Howard, Don. "Holism, Separability, and the Metaphysical Implications of the Bell Experiments." In *Philosophical Consequences of Quantum Theory: Reflections on Bell's Theorem*, edited by James Cushing and Ernan McMullin. Notre Dame, Indiana: University of Notre Dame Press, 1989.

Inwagen, Peter van. "Being, Existence, and Ontological Commitment." In *Metametaphysics: New Essays on the Foundations of Ontology*, edited by David Chalmers, David Manley, and Ryan Wasserman. Oxford University Press, 2009.

Inwagen, Peter van. *Existence: Essays in Ontology*. Cambridge University Press, 2014.

Karras, Tero, Samuli Laine, and Timo Aila. "A Style-Based Generator Architecture for Generative Adversarial Networks." *ArXiv:1812.04948 [Cs, Stat]*, December 12, 2018. http://arxiv.org/abs/1812.04948.

Korman, Daniel Z., and Dana Zemack. *Objects: Nothing out of the Ordinary*. First edition. Oxford, United Kingdom ; New York, NY: Oxford University Press, 2015.

Ladyman, James, and Don Ross. *Every Thing Must Go: Metaphysics Naturalized*. Oxford ; New York: Oxford University Press, 2007.

Lewis, David K. *Parts of Classes*. Oxford, UK ; Cambridge, Mass., USA: B. Blackwell, 1991.

Maudlin, Tim. "Part and Whole in Quantum Mechanics." In *Interpreting Bodies*, edited by Elena Castellani, 46–60. Princeton University Press, 1998.

McDonough, Jeffrey K. "Berkeley on Ordinary Objects." In *The Bloomsbury Companion to Berkeley*, edited by Bertil Belfrage and Richard Brook, 14. Bloomsbury Academic, 2017.

Merricks, Trenton. *Objects and Persons*. New York: Oxford University Press, 2001.

Oppenheim, Paul, and Hilary Putnam. "Unity of Science as a Working Hypothesis," 1958.

Petrovici, Mihai Alexandru. *Form Versus Function: Theory and Models for Neuronal Substrates*. Springer Theses. Cham: Springer International Publishing, 2016.

Quine, W. V. *Ontological Relativity: And Other Essays*. The John Dewey Essays in Philosophy, no. 1. New York: Columbia University Press, 1969.

Radford, Alec, Luke Metz, and Soumith Chintala. "Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks." *ArXiv Preprint ArXiv:1511.06434*, 2015. https://arxiv.org/abs/1511.06434.

Rao, Rajesh PN, and Dana H. Ballard. "Predictive Coding in the Visual Cortex: A Functional Interpretation of Some Extra-Classical Receptive-Field Effects." *Nature Neuroscience* 2, no. 1 (1999): 79–87.

Rescher, Nicholas. *Process Philosophy: A Survey of Basic Issues*. University of Pittsburgh Pre, 2000.

Rosenberg, Alex. *The Atheist's Guide to Reality: Enjoying Life without Illusions*. WW Norton & Company, 2011.

Sacramento, João, Rui Ponte Costa, Yoshua Bengio, and Walter Senn. "Dendritic Cortical Microcircuits Approximate the Backpropagation Algorithm," n.d., 12.

Sacramento, João, Rui Ponte Costa, Yoshua Bengio, and Walter Senn. "Dendritic Error Backpropagation in Deep Cortical Microcircuits." *ArXiv:1801.00062 [Cs, q-Bio]*, December 29, 2017. http://arxiv.org/abs/1801.00062.

Scellier, Benjamin, and Yoshua Bengio. "Equilibrium Propagation: Bridging the Gap between Energy-Based Models and Backpropagation." *Frontiers in Computational Neuroscience* 11 (May 4, 2017).

Schrödinger, E. "Die gegenwärtige Situation in der Quantenmechanik." *Naturwissenschaften* 23, no. 48 (November 1, 1935): 807–12.

Sellars, Wilfrid. *Science, Perception, and Reality*, Ridgeview Publishing Company, 1963.

Smith, Brian Cantwell. *On the Origin of Objects*. Cambridge, MA: MIT Press, 1996.

Smith, Brian Cantwell. "Reply to Dennett." In *Philosophy of Mental Representation*, edited by Hugh Clapin, 237–65, 2002.

Spratling, M. W. "Image Segmentation Using a Sparse Coding Model of Cortical Area V1." *IEEE Transactions on Image Processing* 22, no. 4 (April 2013): 1631–43.

Stebbing, Lizzie Susan. *Philosophy and the Physicists*. New York: Dover Publications, 1937.

Strawson, Peter Frederick. *Individuals: An Essay in Descriptive Metaphysics*. London: Methuen, 1959.

Stroud, Barry. "Transcendental Arguments." *The Journal of Philosophy* 65, no. 9 (1968): 241–56.

Swanson, Link R. "The Predictive Processing Paradigm Has Roots in Kant." *Frontiers in Systems Neuroscience* 10 (October 10, 2016).

Thomasson, Amie L. "Ontological Minimalism." *American Philosophical Quarterly* 38, no. 4 (2001): 319–31.

Wiese, Wanja, and Thomas Metzinger. "Vanilla PP for Philosophers: A Primer on Predictive Processing," 2017, 18.

Wolpert, David H, and William G Macready. "No Free Lunch Theorems for Optimization." *IEEE Transactions on Evolutionary Computation* 1, no. 1 (1997): 67–82.

Zhang, Hui, Jason E. Fritts, and Sally A. Goldman. "Image Segmentation Evaluation: A Survey of Unsupervised Methods." *Computer Vision and Image Understanding* 110, no. 2 (May 1, 2008): 260–80.