CHAPTER 1

Cognitive Kinds

O the mind, mind has mountains; cliffs of fall
Frightful, sheer, no-man-fathomed.

– Gerard Manley Hopkins, "No worst, there is none. Pitched past pitch of grief."

Is there no way out of the mind?

– Sylvia Plath, "Apprehensions"

1.1 Introduction

What is the landscape of the mind? That is the question I aim to tackle in this book. This is an inquiry into the basic components of our mental makeup: What kinds of objects, states, capacities, events, processes, and other entities constitute the stuff of our mental life? As the book's title indicates, the scope is not the mental in general, but the cognitive realm in particular, which I take to be a subset of the mental or psychological realm. Although I will not attempt to demarcate the limits of the cognitive in detail, in what follows, I will attempt to say what characterizes cognitive phenomena, as opposed to other aspects of the mind and brain, later in this chapter (see Section 1.5, as well as Section 2.6). The inquiry is grounded partly in metaphysics and ontology, the philosophical investigation of the building blocks of the universe, and partly in the sciences, empirical research into the workings of the human mind. Since this is a book written by a philosopher, the latter is represented not in the form of original research but by means of distillations of recent empirical work on various mental items and an attempt to synthesize empirical work from different disciplines and subdisciplines. Integrating this empirical work with philosophical argumentation requires paying attention to the relevant literature in cognitive science, including psychology in its various branches (cognitive, developmental, social, and so on), linguistics, neuroscience, computer science, and related disciplines. Given the voluminous amount of work in these areas, it may seem presumptuous to take it all in, and I certainly do not aim to give a comprehensive account of the mental landscape. Instead, I plan to focus on a small number of paradigmatic cases. Of course, this type of integrative project also requires the careful philosophical work of making distinctions, clarifying concepts, and justifying claims with arguments. In this introductory chapter, I intend to lay out some of the philosophical groundwork that supports the argumentation that follows in later chapters. In particular, I plan to spell out the approach to ontology that I intend to take, and specifically the account of categories and kinds that I will adopt, which is naturalist, non-reductionist, and realist (as I will go on to explain).

Inevitably, when one investigates the mind these days, the brain is never far behind. Some would say that the entities constituting the mind are none other than those that comprise the brain, and that we are well on our way to discovering what these are. But despite the fact that there is indeed an intimate connection between psychological and neural entities, I will try to provide reasons for thinking that they are not one and the same and that the categories that pertain to one may not apply to the other. Though the focus will be on mental or psychological entities, their connections and relations to neural entities will often be invoked. To anticipate somewhat, one of the main themes of this book is that there is not always an identity – whether type or token – between psychological and neural constructs, and furthermore, that the validity of a psychological construct does not reside in its coincidence with a neural structure, mechanism, or process. In the neurosciences, there is currently considerable debate and a notable absence of consensus about how mental and neural entities relate to one another. Neuroscientists run the gamut, from those who advocate extreme reductionist positions that posit a "grandmother cell" (see Gross 2002) or a "Jennifer Aniston neuron" (Quian Quiroga, Reddy, Kreiman, et al. 2005) and locate cognitive functions (even particular concepts) in specific brain regions or populations of neurons, to those who preach anti-reductionism and excoriate "blobology," the alleged identification of areas of neural activation with particular psychological capacities, primarily based on regions identified by neuroimaging technology (Poldrack 2012). In subsequent chapters, I will try to provide reasons for thinking that though we are finding and will continue to find many significant correlations between brain structure and cognitive or psychological function, we should not expect a wholesale identification of one with the other. Indeed, I will argue that we will not always be able to identify psychological functions with neural activity, whether or not this activity is localized in specific neural structures. As I mentioned in the previous paragraph, the emphasis will be on *cognitive* ontology rather than psychological ontology more broadly. All the case studies to be discussed involve cognition in some way, as opposed to affective, perceptual, sensory, or experiential aspects of mentation. The aim is not to give an exhaustive catalogue of the contents of the mind (if that were even possible) but rather to focus on a range of significant examples of categories that involve cognition, examine the case for admitting each into our ontology, and draw some general conclusions about the kinds of entities that we should posit in cognitive science and on the grounds for doing so. After this first programmatic chapter, each of the rest of the book's chapters tackles one or a small number of candidates.

1.2 Naturalism about Kinds

In investigating mental objects, states, capacities, events, processes, and other entities, we are usually investigating types not tokens, that is, not unique particulars, but types or *kinds* of them. Specifically, we are interested in which of these types or kinds are real or "natural," or in standard philosophical parlance: *natural kinds*. Many contemporary discussions of natural kinds base their notion of kinds on the essentialist account first sketched out by Putnam (1975) and Kripke (1980). Instead, I will anchor the account of kinds that I will be deploying throughout this book in a nineteenth-century tradition that is more closely aligned with a naturalist philosophical outlook. According to the naturalist tradition that I will be tapping into, empirical science is our best guide to the kinds that exist in nature, rather than a priori considerations from metaphysics or philosophy of language. This attitude originates with the discussion of scientific classification that is prominently represented in the works of Whewell and Mill, and indeed in their mutual influence. Though Mill is often credited with initiating the discussion of natural kinds (or just plain "kinds," as he called them) in modern philosophy, even a casual reader of Mill's *A System of Logic* (1843/1882) cannot help but notice the considerable debt to Whewell's *Philosophy of the Inductive Sciences* (1840/1847). Despite significant differences in their overall philosophical positions, Whewell being a neo-Kantian rationalist and Mill a staunch empiricist, there is much that they agree on when it comes to kinds. Whewell and Mill both regard science as the guide to uncovering kinds in nature and think that scientific taxonomy aims at discovering kinds. Moreover, they are both concerned with the rational grounds for scientific classification and are keen to understand the differences between "natural" and "arbitrary" scientific classification schemes. They both see kinds as the basis for inductive inference and regard science's search for kinds as a quest to come up with categories that would lend themselves to empirical generalizations and natural laws. As Whewell writes: "The object of a scientific Classification is to enable us to enunciate scientific truths: we must therefore classify according to those resemblances of objects … which bring to light such truths" (1840/1847, 486). Whewell also thinks that classification must not be based on any resemblances whatsoever but on what he calls "natural affinity," which requires us to classify things on the basis of properties that generally cooccur with other properties (1840/1847, 542). Moreover, he repeatedly states that "the great rule of all classification" is that "the classification must serve to assert general propositions" (1840/1847, 495). Mill endorses this emphasis on "general propositions" or "general assertions" and goes on to say that "the very first principle of natural classification is that of forming the classes so that the objects composing each may have the greatest number of properties in common" (1843/1882, 879). Hence, for both Whewell and Mill, the aim of scientific classification is to group things together based on shared cooccurring properties, so that the categories that result enable us to make valid scientific generalizations.

While the naturalist tradition that originates with Whewell and Mill provides the main philosophical inspiration for the account of kinds that I will be operating with in this book, there is one respect in which I will part company with this older tradition. These philosophers are not very clear when it comes to the metaphysics of kinds. They seem to think that uniformities in nature are the basis for successful scientific generalization and inference, but they do not fully explicate the nature of these uniformities. Venn (1889/1907) criticizes Mill for distinguishing between two kinds of uniformity in nature: uniformities of sequence (which are causal) and uniformities of coexistence (which are brute). By contrast, Venn thinks that many of the uniformities of coexistence identified by Mill are actually causal in nature (though he does not think that *all* uniformities in nature are causally based). Still, he holds that uniformities are what enable us to use natural kinds in inductive inference in science. According to Venn (1889/1907, 94), uniformity "is the objective counterpart or foundation of inferribility …." Inductive inferences are based on uniformities and are therefore dependent on the existence of kinds in nature, which reflect these uniformities. Thus far, I agree with Whewell, Mill, and Venn. But by contrast with them, I will assume that uniformities in nature are due to regular and stable connections between causes and effects, and that these *causal relations* are the metaphysical bases of scientific induction and epistemic practices.[1] This assumption is also shared by many contemporary naturalist philosophers. It is prominent in Boyd's account of natural kinds and it is exemplified in what he calls the "accommodation thesis": "Kinds useful for induction or explanation must always 'cut the world at its joints' in this sense: successful induction and explanation always require that we accommodate our categories to the causal structure of the world" (1991, 139). Boyd also speaks of "the accommodation of inferential practices to relevant causal structures" (2000, 56).[2] This is also a central feature of Kornblith's (1993, 35) account of natural kinds: "It is precisely because the world has the causal structure required for the existence of natural kinds that inductive knowledge is even possible." This link between the epistemology of categories and the ontology of kinds is characteristic of a naturalist attitude to metaphysics, which holds that our metaphysical inquiries should be guided by our best epistemic practices as exemplified in the considered classification schemes of our best scientific theories. Among at least some contemporary naturalist philosophers, the causal structure of the world is the ontological basis for the successful epistemic practices of science.

Contemporary naturalist philosophers think that the causal uniformities in nature, even those discovered by the basic sciences, are rarely if ever ironclad or exceptionless, and this implies that the properties associated with natural kinds are loosely clustered rather than invariably associated with one another. Moreover, as I have already mentioned, the properties that cluster in kinds are not just sets of properties that happen to cling together, since they are associated as a result of causality. Accordingly, rather than view kinds as mere clusters of properties, I have proposed that they be conceived as "nodes in causal networks" (Khalidi 2013; 2018). According to this "simple causal theory" of natural kinds (cf. Craver 2009), certain properties or conjunctions of properties that are causally connected with others in systematic ways can be considered natural kinds. Sometimes we identify the kind with just one of the properties in a causal chain or network,

[1] There may be some uniformities in nature that are brute and not causally based, particularly at the most fundamental level. But I will assume that these are not at issue in a discussion of cognitive ontology. For further justification, see Khalidi (2013; 2018).

[2] Elsewhere in the same paper, Boyd emphasizes the ways in which natural kinds are "practice-dependent" and relative to human interests, and it is not easy to reconcile this attitude with his accommodation thesis. On the view that I favor, human interests serve only to select certain causal structures and processes to focus on, they do not somehow shape or modify them (except in cases in which humans are themselves part of the causal process – see Section 1.5).

but at other times we draw a wider circle among a number of them and consider that set of properties to be the kind. Either way, we are identifying properties that are causally conjoined to others, rather than mere clusters of properties. This causal account of kinds is somewhat less restrictive than that proposed by Boyd, who considers kinds to be property clusters that are held in homeostasis by causal mechanisms – though he sometimes relaxes these conditions and gestures toward something like a simple causal account. Thus, the simple causal account is distinct from a strict version of Boyd's account, which requires a specific causal mechanism to keep the cluster of properties in equilibrium (or homeostasis). I have questioned the strict version on two grounds. First, in many cases, there is nothing that can properly be called a causal mechanism that holds the properties together – they may instead be held together functionally or relationally, as we shall see in later chapters. Second, the properties involved are not always in a state of equilibrium – they may be repeatedly instantiated through the action of independent causes.[3] A simple causal theory of kinds can also be usefully distinguished from an essentialist one, at least on many versions of essentialism. Though essentialists also tend to think that natural kinds are discoverable by science, they usually place additional conditions on natural kinds, which I think are at odds with scientific taxonomy. There are four ways in which this account of kinds differs from many essentialist ones. First, the properties that are associated with each kind are causally linked, but they can consist in a loose cluster rather than a set of properties that are both necessary and sufficient for kind membership. Second, the causal properties may be functional or relational rather than intrinsic. Third, the properties involved do not have to be microstructural, as some essentialist philosophers tend to insist. Fourth, the simple causal theory does not claim that these properties are associated with the kind in question across possible worlds or with modal necessity, as essentialists usually hold.

Another significant point of agreement in the naturalist tradition that stems from Whewell, Mill, and Venn is that natural classification schemes and the kinds that they identify can be found across the sciences, including the human sciences. These philosophers tend to see considerable continuity from chemistry to mineralogy to biology to psychology and the social sciences, especially when it comes to the importance and feasibility of uncovering kinds. This attitude seems less prevalent among contemporary

[3] These claims are further justified in Khalidi (2013; 2018). I have also proposed that natural kinds can be represented by means of directed causal graphs. Although I have not worked out this proposal in detail, in such representations, natural kinds correspond to highly connected vertices in directed causal graphs.

philosophers, at least some of whom think that it is a truism that natural kinds pertain to the natural sciences. Hence, it may appear oxymoronic to talk about natural kinds in the cognitive sciences. Given that the terminology of "natural kinds" is misleading, especially in the context of the human sciences, I will be talking mainly of "kinds" or "real kinds" instead of "natural kinds," especially given that the very existence of the expression "natural kind" seems to be a historical accident. As Hacking (1991) has pointed out, the terminology of "kinds," which Whewell and Mill used, gave way to "natural kinds" as a result of the writings of Venn. But Venn seems to have taken himself to be using Mill's expression, since he credits him with introducing the term – despite the fact that Mill apparently never used it. Venn (1889/1907, 84) writes: "he [i.e. Mill] introduced the technical term of 'natural kinds' to express such classes as these." It is unclear whether Venn simply misremembered Mill's terminology or whether he deliberately modified it. Either way, we are now saddled with an unfortunate expression, which is misleading on at least two counts. The first reason that the expression "natural kind" is deceptive is that it tends to set up a misguided contrast between the natural and *artificial*. In many scientific domains, there are strong candidates for kinds that have the "trail of the human serpent" over them and may reasonably be considered artificial (especially in the Anthropocene era). Whether we are dealing with synthetic chemicals, genetically engineered organisms, or artificially intelligent systems, scientists now study a range of entities that are the result of human intervention (if not wholesale invention), yet apparently no less real or objective than their supposedly "natural" counterparts. But the terminology of "natural kinds" would encourage us to dismiss the kinds to which these entities belong. The second reason the expression is misleading (which is more important for these purposes) is because the adjective "natural" suggests an affiliation with the natural rather than the social sciences, and it threatens to sideline categories that have a social or human dimension. When it comes to the cognitive sciences, which straddle the biological and psychological sciences, this is especially pernicious, since it tends to privilege the former over the latter, perhaps suggesting that neural kinds are more objective than psychological ones.

Here, it may be objected that the philosophical apparatus of real kinds may not be the right lens through which to view cognitive science. It may be thought that kinds are more at home in sciences like botany or mineralogy, where the paradigmatic individuals are well-defined concrete particulars (individual plants, mineral samples), with clear spatiotemporal boundaries. In cognitive science, though there are some fairly neat individuals such as human persons (and other creatures), which are often classified into kinds (e.g. *schizophrenic*, *bilingual*), the individuals can also be cognitive modules, cognitive capacities, mental states, mental processes, and other entities, so it may not be as useful to think of such entities as belonging to kinds. I would reply simply by stressing the indispensability of taxonomy to any scientific discipline or subdiscipline. Whenever we theorize about any domain, it is inevitable that we classify the items that populate that domain and that we do so in nonarbitrary ways. Classification, in turn, presupposes dividing a domain of entities into types or kinds. Moreover, as I will try to show, although some of the items classified in cognitive science are not best thought of as individuals, but states, capacities, events, processes, and so on, they are also divisible into kinds. Hence, there is no need to think of classification as pertaining exclusively to a domain in which concrete particulars with well-defined spatial boundaries are the main items of interest.

This brief sketch of a naturalist theory of kinds and its underlying metaphysics will have to suffice for now. More details will emerge as we survey a number of candidates for cognitive kinds in subsequent chapters.

1.3 Ontological Matters

In recent philosophical and scientific discussions of cognitive ontology, it is common to read that "ontology" is used differently by philosophers and others, namely psychologists, neuroscientists, and perhaps most prominently, computer scientists. I believe that this claim is not wholly justified. There are perhaps some differences in emphasis and nuance in the usage of these disciplines, but this is not a case of sheer polysemy. The main difference may be that computer scientists (in particular) are interested in how domains are taxonomized without great regard to how they *ought* to be taxonomized, and without a commitment to the domain's actually containing the entities that are posited by the taxonomic or classificatory system. Philosophers, on the other hand, tend to be interested in the *ought* and in the underlying structure of reality. As emphasized in Section 1.2, naturalist philosophers tend to think that our current, mature, scientific taxonomic systems are our best (defeasible) guides to that underlying structure. In other words, they derive an *ought* from an *is*.[4] This is warranted on the

[4] This is a stark and provocative way of putting it. For a more nuanced account of the relationship between scientific practice and philosophical theory, see Khalidi (2013), where I lean on the notion of "reflective equilibrium," first introduced by Goodman (1954/1979).

assumption that science aims at discerning that structure. In later chapters, we will encounter challenges to that assumption, on the grounds that some investigators are not just aiming to discern the causal structure of the world, since their inquiries are shaped by non-epistemic norms (especially in areas like psychiatry). I will put such concerns to the side for now, and will take them up in some subsequent chapters (but see also Section 1.5).

If we are naturalists, then talk of "ontology" is closely related to talk of "taxonomy" or "classification" – provided we think that science aims primarily at classifying entities in such a way as to discern the causal structure of the world, and is guided in doing so by epistemic goals. When viewed thus, there does not seem to be an equivocation or ambiguity in the use of the term "ontology" and related expressions. If we bear in mind that "ontology" should not be used as a synonym for "taxonomy" or "classification scheme," but rather to denote the metaphysical structure that is described by a taxonomic system or classificatory scheme, then some of the differences in usage can be cleared up. This caveat is also relevant to the use of terms like "kind" and "category." These two terms (and related ones) are often used interchangeably, by philosophers and cognitive scientists alike, but I propose to distinguish them, as follows. A *kind* should be understood to be an entity in the world, which can be conceived of as a collection of particulars or set of entities (nominalist reading), or an abstraction, such as a universal that is immanent in particulars (realist reading). Meanwhile, a *category* pertains to our conceptual, theoretical, or linguistic framework and practices; it is the concept of a kind.[5] In other words, a kind pertains to ontology whereas a category pertains to taxonomy. Here again, the two notions are closely related, since (on a naturalist understanding) the aim of scientific inquiry is to devise categories that correspond to all and only the kinds.

Once we distinguish ontology and taxonomy, along with kinds and categories, we should take care not to embrace a view that has been derided as the "third dogma of empiricism" (Davidson 1973). According to this dogma, we can somehow confront our kinds with our categories directly to determine whether they are in alignment, as we might compare a map of the landscape with the terrain itself for accuracy. The problem with this way of thinking is that we have no access to the "terrain" that is not mediated by our "map" (which is why the cartographic analogy is so misleading). We access the world via the categories of our taxonomies and hence, we cannot step outside of them to see how well they align with the

[5] To mark this distinction, I will generally italicize *kinds* and put CONCEPTS in small caps.

world itself. But that does not mean that we have no way of determining whether and how well our categories delineate the kinds. Since we devise these categories to describe the world, we can determine how well they enable us to generalize, explain, predict, and so on. Depending on their efficacy in fulfilling our epistemic goals, we infer that they have or have not latched on to the causal structure of the world, in line with the naturalist picture outlined in Section 1.2. This view has been articulated lucidly by Child (2001, 38), who writes: "in classifying things by reference to their causal powers, or their causally significant composition, we classify things in ways that reveal the way the world works." As long as human inquiry is able to achieve this goal, the specification of an ontology is not beyond our reach.

I argued in Section 1.2 that classification and the identification of kinds is based on identifying properties that are associated with other properties, and these properties are so identified because they are causally related. But what about those properties themselves? How do we identify the most basic properties in our ontology, and might we have settled on a different system of kinds if we had started with a different set of properties? This is an old philosophical conundrum and I cannot pretend to give a satisfactory answer to the question in the scope of this book (see e.g. Goodman 1954/1979; Lewis 1983). It is true that real kinds are grounded in shared properties and that these properties may be considered the unjustified posits upon which the whole theoretical edifice is built. If the properties to which we humans are attuned are just reflections of our parochial perceptual and cognitive abilities and do not reveal real features of the universe, then you might say that we have no reason to believe that the kinds that we identify expose the real joints in nature. In cognitive science, such properties might include basic behavioral ones involving motion, force, space, and time (e.g. eye movements, button presses, looking times, reaction times) or more abstract intentional ones (e.g. expressed preferences, discrimination between stimuli). But there may be a way of overcoming certain skeptical doubts about these baseline properties. For there is an indirect vindication of our choice of baseline properties in the identification of kinds that enable us to make generalizations, which in turn help us to explain and predict the entities in question. As I argued in the previous section, these epistemic desiderata are themselves causally based, so the choice of properties is ultimately upheld by our ability to use them to understand the causal structure of the world. Unless causality is itself an illusion, or a mere reflection of our inadequate and distorted perceptual and cognitive endowments, our choice of properties in cognitive science is