CHAPTER 2

Concepts

I have a young conception in my brain;
Be you my time to bring it to some shape.
– William Shakespeare, *Troilus and Cressida*

Every new concept first comes to the mind in a judgment.
– C. S. Peirce, "Belief and Judgment"

2.1 Introduction

Concepts are often considered the most basic element in our cognitive ontology; they are frequently regarded as the building blocks of the mind or the vehicles of thought. The past few decades have seen a surge of work when it comes to concepts, both theoretical and empirical. Once the exclusive province of philosophers, concepts have become the stomping ground of cognitive scientists over the past several decades. There are currently a number of different research programs investigating concepts empirically using a variety of methods. Even though this is seldom made explicit, I would argue that they do not always address the same questions. There are at least five different questions that are commonly at issue in recent theoretical and empirical work on concepts. First, what types of entities are concepts? Can they be *identified* with mental entities, neural entities, both, or neither? Are they concrete particulars or abstract entities of a certain kind? Second, there is a question about concept *individuation* or grounding: In virtue of what does a concept have the content that it does? In other words, what gives a concept its identity conditions, or what makes something the concept of APPLE as opposed to ORANGE? Third, researchers are interested in concept *acquisition*: How are concepts acquired in ontogeny? Are some concepts innate, or are they all learned, and by what processes are concepts learned? (There is also some research concerning the

https://doi.org/10.1017/9781009223645.003 Published online by Cambridge University Press

acquisition of concepts in phylogeny, though for obvious reasons there is less empirical work relevant to that question.) Fourth, there is a question about concept *possession*: What is it for a thinker to have a concept, or what determines whether a subject has acquired a concept? Finally, there is the issue of concept *activation*: What is it for a concept to be activated in the mind on a particular occasion (which is roughly the same as conceptual retrieval or processing)? What goes on in the mind or brain of thinkers when they entertain a particular concept? To anticipate, I will argue in due course that some research programs are more focused on some of these questions explananda than others. Hence, they are sometimes working at cross purposes and are not attempting to answer the same questions or explain the same phenomena.

In addition to the fact that there are at least five different research questions or explananda that the research on concepts addresses, there are a number of theoretical controversies that split researchers on concepts into various camps. These theoretical controversies do not map neatly onto the explananda mentioned above, in the sense that some of the theories provide an answer to more than one of the following questions or attempt to explain more than one of these phenomena (without always clearly distinguishing them). What follows is not an exhaustive list of debates about concepts, but these theoretical disputes are among the most prominent ones and they are the ones that will figure in the discussion to follow:

1) ***Externalism vs. Internalism (Individualism)***: An "externalist" view of concepts holds that concepts are individuated by determinants that are external to the mind of the agent, while an "internalist" (or "individualist") position attends to the subject's perspective on the world in individuating concepts.

2) ***Modal (Sensorimotor) vs. Amodal Theories***: This is a distinction, which is grounded in some traditional philosophical debates as well as in the recent empirical literature, between accounts of concepts that are based in our sensory modalities or sensorimotor abilities and those that are amodal (or trans-modal). The former regard concepts, whether concrete or abstract, to have their contents as a result of their connections to sensory, affective, and motor functions or processes, whereas the latter do not think that sensory, affective, and motor abilities can account for the contents of concepts.

3) ***Definitional (Classical) vs. Prototype vs. Theory Theories***: Another debate is that between those who take the content of concepts to be supplied by a definition or consider them to be structured in terms of

https://doi.org/10.1017/9781009223645.003 Published online by Cambridge University Press

necessary and sufficient conditions, and those who deny this, judging the content of a concept to consist of a prototype of weighted features with a family resemblance structure. There are also other accounts of conceptual content or structure, most notably, those accounts that take concepts to be embedded in theories and inextricable from them.

4) ***Holism vs. Atomism***: This debate concerns whether the content of a concept depends in systematic ways on the contents of the other concepts in a conceptual repertoire ("holism"), or whether each concept has its content independently of others ("atomism"). One can also distinguish an intermediate position that considers the content of each concept to depend only on *some* other concepts ("molecularism").

5) ***Response-Dependence vs. -Independence***: This is a dispute between those who consider the identity conditions for concepts to be dependent ultimately on human responses and judgments, and those who do not hold them to be response-dependent in this way. If concepts are response-dependent, that means that they are phenomena whose very identity is determined by the judgment of observers or interpreters (this is the case for an "interpretive" or "ascriptive" view of concepts). If they are not, then the content of a concept or its individuation conditions are not dependent on the response of an interpreter or the judgment of the community.

6) ***Minimalism vs. Intellectualism (Maximalism)***: There is a distinction that is sometimes made between "intellectualist" theories of concepts, which consider that language and higher cognitive abilities are required for the possession of concepts, and "minimalist" theories of concepts, which hold that they merely involve discriminatory abilities or sensitivities along with certain combinatorial abilities. The latter account of concepts is supposed to countenance the possession of concepts by prelinguistic infants and nonhuman animals, whereas the former is thought to restrict concept possession to language-using creatures.

These would seem to be the most prominent points of disagreement among those researchers – philosophers, linguists, psychologists, neuroscientists, and others – who are engaged in inquiring into the nature and structure of concepts. In what follows, I will be touching on each of these controversies to various degrees and will in some cases propose ways of reconciling the different approaches or at least rendering them compatible. To illustrate the point that these theoretical approaches do not correspond neatly to the earlier research questions outlined earlier, consider the first theoretical dispute between externalists and internalists. This is primarily a disagreement concerning the question about conceptual individuation, since these

https://doi.org/10.1017/9781009223645.003 Published online by Cambridge University Press

theorists give different answers to the question as to what gives concepts their contents or semantic values. But these theorists also tend to differ when it comes to the second question concerning acquisition, since many externalists think that concepts are acquired by thinkers (roughly) when they are causally connected in some way to the referents of those concepts, whereas internalists tend to think that thinkers can acquire concepts by relating them to other concepts in the right ways. This also leads to a difference of opinion about conditions on concept possession. Similar points apply to some of the other theoretical debates mentioned in the previous paragraph, which have repercussions for more than one of the questions mentioned: identification, individuation, acquisition, possession, and activation.

The rigorous empirical study of concepts has only been around for roughly half a century, but scientists have devised various methodological paradigms to ascertain concept activation, as well as concept possession and acquisition. There is a wide range of methods and tasks used, many of them involving responding to verbal stimuli and images that represent, illustrate, or are otherwise associated with particular concepts. The concepts investigated are predominantly those denoted by common nouns for concrete objects (e.g. "apple," "chair," "fruit," "furniture"), though concepts denoted by verbs are occasionally also studied. The tasks usually involve categorization (including assent or dissent from statements concerning categorization, e.g. "an apple is a fruit"), recognition, discrimination, and inference. When experimenters rely on verbal and behavioral responses, these responses are sometimes elicited under time constraints or while measuring reaction times. In some cases, there are no such constraints and participants answer at their leisure and are asked to justify their responses. Finally, in addition to verbal and behavioral responses, studies sometimes involve neural ones, measured either by the blood-oxygen-level-dependent (BOLD) signal in a fMRI scan or the electrophysiological reading indicated by an electroencephalogram (EEG). Though there are many experimental paradigms that are not covered by this brief overview, these are some of the typical methods used in empirical research on concepts, and we will encounter them again in what follows. In Section 2.2, I will present some research on concepts in cognitive neuroscience, and introduce two theoretical approaches that have emerged from this work, which relies primarily on neural evidence. Then, in Section 2.3, I will survey some of the main results in research on concepts in cognitive psychology, which depends on evidence from behavioral and verbal responses, as well as the main theoretical constructs. After that, in Section 2.4, I will present a functional theory of concepts that has some affinity with one

https://doi.org/10.1017/9781009223645.003 Published online by Cambridge University Press

theoretical approach in cognitive psychology, arguing that other theories and taxonomic categories in psychology and neuroscience may be tracking different cognitive processes and identifying different kinds. In Section 2.5, I will consider and reply to a number of objections to this theory of concepts, before concluding in Section 2.6.

2.2 Empirical Accounts: Cognitive Neuroscience

Most cognitive scientists identify concepts with types of mental representation, and for many, mental representations are supposed to be implemented in certain types of neural entities or processes (e.g. populations of neurons, patterns of neural activations), though few if any cognitive scientists would claim to know exactly how they are so implemented at present. Many researchers on concepts, especially in cognitive neuroscience, are therefore interested in isolating the neural correlates of concept activation, paving the way for a reductionist account of what it is for a thinker to entertain a concept on a particular occasion. Some of these theories also contain or lend themselves to a particular account of how concepts get their contents, or what it is about a particular pattern of neural activity that constitutes the activation of a particular concept. They are hence theories of individuation or grounding as well as activation. They can also be seen to give an account of possession, though this is not usually their focus, as I will go on to argue. In this section, I will present what is currently the dominant theory of concept activation and individuation in cognitive neuroscience, which closely associates conceptual representations with perceptual ones. (For brevity, I will sometimes talk exclusively of perceptual or sensory systems, but motor and affective systems should also be understood to be implicated in the activation of at least some concepts, on this view.) After outlining what I take to be some of the main problems with this theory, I will then sketch an alternative picture that has been proposed by some cognitive neuroscientists, which problematizes the search for concepts in neural infrastructure.

On a modal theory of concepts the deployment of a concept on a particular occasion consists in the occurrence of patterns of neural activation that reactivate states in modality-specific systems in the brain, namely perceptual, motor, and affective systems. On one version of this theory, the activation of a concept consists in large part of the "re-enactment" of sensorimotor perceptual representations associated with that particular concept, where a reenactment "partially reproduces experienced states" (Barsalou, Simmons, Barbey, et al. 2003, 88). Since the mere reenactment

https://doi.org/10.1017/9781009223645.003 Published online by Cambridge University Press

of perceptual features would not seem to account for full-fledged conceptual abilities, including the ability to categorize and infer, many accounts also invoke additional neural processes. On most modal views, it is not enough for these neural populations to be coactivated to be bound together in a single concept, they have to be conjoined in some way, and on some, conjunctive neurons are responsible for binding and reactivating neuronal populations responsible for encoding perceptual features. Hence, the reenactment consists not just in the reactivation of perceptual representations but also in the activation of "cross-modal" neurons in "convergence zones," which are also known as "association areas" or "hubs" in the brain (Barsalou 2016, 1132). These convergence zones are supposed to be responsible for binding the percepts together and are posited to integrate information across modalities. Thus, among modal theorists, many claim that all that is required for the representation of even abstract concepts is perceptual representation along with certain conjunctive operations carried out by cross-modal neural systems, and they think of "abstract conceptual representations as high-level conjunctions rather than amodal symbols" (Binder 2016, 1098). While some modal theorists occasionally invoke "amodal" representations, this would seem to undermine the theory's most basic tenet that conceptual representation is fundamentally modally grounded. Moreover, as some have pointed out, "amodal" as used by modal theorists should really be taken to mean "multimodal" (a term often used interchangeably with "crossmodal" and "transmodal," or occasionally "supramodal"), and is used to describe representations that somehow combine or bind representational content from multiple sensory modalities (Barsalou 2016, 1126; Binder 2016, 1098; Spunt & Lieberman 2012; cf. Kemmerer 2019, 41 n.1). This account is supposed to apply to abstract as well as concrete concepts, and indeed erase the distinction between the two types of concept (Barsalou, Dutriaux, & Scheepers 2018).

Some evidence for modal theories of concepts comes from experiments that exclusively rely on behvioral techniques and measures, such as reaction times, error rates, and frequency of listing certain features. For instance, as reviewed by Barsalou, Simmons, Barbey, et al. (2003), when participants are asked to list features associated with a concept, features are reported less if they are typically occluded perceptually (e.g. participants produce "roots" less often for "lawn" than for "rolled-up lawn"), and when participants are asked to verify sensory features associated with certain concepts they are slower if they had been asked to verify a feature from a different modality on a preceding trial (e.g. verifying "loud" for blender is faster after previously verifying "rustling" for leaves than after verifying "tart" for cranberries).

The experimental design presupposes that if conceptual representations are modality-specific, switching from one modality to another should slow verification. But since these are at best indirect indications of the underlying structure or neural organization of conceptual representations, more compelling evidence for modal theories of concepts comes from recently developed experimental techniques in cognitive neuroscience. Though the tasks in these experiments are primarily standard behavioral tasks involving categorization or recognition, the main methodological development is the use of neuroimaging technology to measure neural activity while participants undertake such tasks. The variety of experimental protocols that have been used in this area of research is difficult to summarize here, but most experiments require experimental participants to perform certain cognitive tasks relating to concepts while undergoing a fMRI scan. In a standard experimental paradigm, participants are simply asked to read individual words in the scanner while their hemodynamic activity is monitored. In one highly cited study, experimenters selected a number of action words associated with face, arm, and leg movements (e.g. "lick," "pick," "kick"). Participants in the experiment were asked to read individual words on a screen while fMRI was used to gauge levels of neural activation in their brains (Hauk, Johnsrude, & Pulvermüller 2004). The hemodynamic response when reading these words was then compared to the response when participants were asked to move their left or right foot, left or right index finger, or tongue. In this instance, experimenters found that the action words differentially activated areas along the motor strip that were either directly adjacent to or overlapped with areas activated while moving the corresponding body parts, thereby providing support for a modal theory of concept representation. In such an experimental paradigm, it is difficult to ensure that subjects are indeed focusing on the conceptual content of the words they are reading and that they are not simultaneously distracted by other thoughts. Indeed, for any given task, it is problematic to determine which precise concepts are being activated, since there is seldom just a single concept at play in any given cognitive task, even under controlled experimental conditions. Thus, to avoid confounds, various attempts have been made to design tasks that would serve to zero in on the concept or concepts of interest in any given condition. In a more recent experimental design, a written word denoting a concept was presented on the screen for several seconds and participants in the scanner were asked to "think deeply" about its meaning in such a way that they could determine subsequently whether it applied to a visual scene (Wilson-Mendenhall, Simmons, Martin, et al. 2013). The visual scene was then presented and participants were asked to judge whether or not the

https://doi.org/10.1017/9781009223645.003 Published online by Cambridge University Press

word applied to that scene. In some trials, the relevant words were not followed by a visual scene, and these "catch" trials supposedly enable researchers to separate the neural activity involved in thinking "deeply" about the concept from that involved in judging whether or not it applies to a visual scene. This way, researchers are supposed to capture the neural activation associated with deploying a concept. But despite the ingenuity of this method, there are various questions that arise even when it comes to this careful attempt to capture the activation associated with a specific concept and separate it from other cognitive processes. Perhaps most importantly, it is far from clear that asking people to explicitly think about or reflect on a concept is the best way of eliciting that concept. If one of the main functions of concepts is categorization, then the process of applying a concept to a visual scene might actually be a more valid way of gauging the neural activation associated with conceptual processing or deployment.[1]

The behavioral and neural evidence for modal theories of concepts or concept activation is compelling, but it raises a number of questions that should lead us to exercise caution. First, as already mentioned, many, if not most, concepts do not seem to be mere concatenations of perceptual features, even concrete concepts like APPLE. At the very least, modal theories need to say far more about how convergence zones enable the representation not just of simple conjunctive relations (e.g. RED and ROUND), but more complex logical and statistical ones (e.g. if RED then probably RIPE), let alone the representation of concepts that do not seem amenable to being constructed simply out of clusters of perceptual features (e.g. RIPE). Even if we allow that conjunctive neurons or convergence zones manage to bind perceptual representations in the logical relationship of conjunction, no explanation is forthcoming of other Boolean and non-Boolean aspects of conceptual structure (e.g. disjunction, if-then, part-whole, set-subset, predication, probabilistic, generic, and so on). But this would require drawing on hitherto unspecified non-perceptual representational resources. In a critique of modal theories, Weiskopf (2007, 174) has argued forcefully that when it comes to the representational properties of at least some concepts, "we will need to appeal to many neural activation patterns

[1] Another problem with this experimental design, as well as many others in this area, is that they rely to some extent on the "method of subtraction," which attempts to isolate the neural correlate of some cognitive process C_1 by starting with some task T, which is thought to involve various cognitive processes, C_1, C_2, … C_n, and then subtracting from the neural correlate of T the neural correlates of the other processes, C_2, … C_n, as revealed in other experiments. But as many researchers have pointed out, this problematically assumes a simple additive relationship between cognitive processes (see e.g. Poldrack & Yarkoni 2016).

beyond those in the perceptual systems; these neural activation patterns are not themselves perceptual representations or copies thereof ...” At least as they stand, modal theories do not seem to account for a panoply of conceptual abilities that transcend perceptual (as well as motor and affective) features and include theoretical information required in making complex inferences about such matters as causality and intention. Moreover, this problem implies that modal theories of concepts do not have an obvious advantage when it comes to giving an account of concept individuation or grounding. One of the main selling points of modal theories would seem to be their ability to provide such an account, simply because conceptual representations can be said to inherit their representational content from perceptual ones, which are presumably grounded by being connected in the appropriate way to sensorimotor receptors. However, given that modal theories owe us an account of “missing” representational content, they must also provide an alternative way of accounting for grounding at least some non-perceptual aspects of conceptual content.[2]

Second, despite the existence of significant empirical results that appear to confirm modal theories of conceptual activation, there is a substantial body of counter-evidence, which has been detailed by the theory’s critics. Some of this evidence consists in the existence of double dissociations among patients with neural lesions. For example, in the color domain, patients with lesions can be impaired in their color knowledge (e.g. “this is the color yellow”) but can nevertheless retain knowledge of the typical colors of objects (e.g. “bananas are typically yellow”). Conversely, lesion patients can have intact color perception but impaired conceptual knowledge of the typical colors of objects (see Mahon & Hickok 2016, 949, and references therein). Numerous examples of such dissociations exist between sensorimotor capacities and conceptual capacities, indicating that it would be too hasty to conclude that conceptual representations are simply reenactments of perceptual representations, or even that they involve such reenactments as a necessary element. There is additional neuropsychological evidence against modal theories from “semantic dementia,” which results in a “modality-general, item-specific” pattern of deficit, with “no apparent interaction between conceptual category and perceptual modality” (for a review and analysis, see McCaffrey 2015a, 343).

Third, modal theories have a problem when it comes to distinguishing the conceptual activation pertaining to a concept proper from activation

[2] Barsaolu, Dutriaux, and Scheepers (2018) seem implicitly to acknowledge this by attempting to draw on the resources of situated cognition to flesh out the representational content of concepts.

https://doi.org/10.1017/9781009223645.003 Published online by Cambridge University Press

associated with a concept but not strictly proprietary to the concept itself. In this vein, Mahon and Hicock (2016, 947) point out that at least two inferences can be drawn from empirical evidence indicating that sensorimotor representations are activated during conceptual processing. The first is that concepts are represented in a sensorimotor format and activation of sensorimotor representations reflects conceptual access. The second is that concepts are represented in an amodal format and that activation spreads from these amodal representations to connected sensorimotor representations, whose activation is evidence of information flow from one system to the other. The latter inference is effectively an alternative hypothesis to explain some of the same evidence that is usually taken to support the modal theory. To rule out this hypothesis, one would have to have a principled way of distinguishing the neural activation constituting activation of the concept proper from neural activation that is associated with the concept but not constitutive of it. This poses both technical and theoretical challenges. Technically, the temporal resolution of many of our current neuroimaging tools may not be sufficiently fine-grained to distinguish the relevant cognitive processes. Conceptual processing in many experimental tasks can take place in the space of a few hundred milliseconds, whereas the temporal resolution of fMRI scans is currently of the order of one thousand milliseconds. Theoretically, it is notoriously difficult to use "functional connectivity," which is a measure of statistical correlations between regions as revealed in fMRI scans, to infer "effective connectivity," which is a measure of how regions actually interact causally. Perhaps more importantly, as I will go on to detail, there is mounting evidence to suggest that there may not be a clear theoretical distinction to be made between the activation associated with the concept itself and at least some types of associated activation. Much recent work on conceptual processing has pointed to the variability in neural activation associated with a given concept, especially as it occurs in different contexts. According to this body of research, there is no "core" of neural activation associated with any given concept across different contexts, including experimental tasks, individuals, languages, and cultures, among others. In the rest of this section, I will briefly outline this body of work and extract its relevance for the attempt to identify the neural correlates of conceptual thought.

Another major trend in recent empirical work on concepts in cognitive neuroscience emphasizes the fluidity, flexibility, and instability associated with conceptual content. The main findings in this body of research concern the inter and intrapersonal variability in neural activation when it comes to tasks involving the deployment of concepts, such as categorization

https://doi.org/10.1017/9781009223645.003 Published online by Cambridge University Press

and recognition. Yee and Thompson-Schill (2016) review a number of neuroimaging studies showing that neural activation varies along a variety of contexts or dimensions when it comes to conceptual tasks: long-term experience, recent experience, concurrent context or ongoing goals, and passage of time as object recognition unfolds. For example, when it comes to long-term experience, in professional musicians, identifying pictures of musical instruments activates auditory association cortex and adjacent areas more than identifying pictures of other objects, but this difference in activation does not appear in musical lay people (Hoenig, Müller, Herrnberger, et al. 2011). In such cases, modal theorists might just say that experts and nonexperts simply have different concepts, but similar results obtain for categorization and recognition tasks when it comes to the other conditions mentioned, such as a person's recent experience. For example, even though many neuroimaging experiments confirm that words relating to actions recruit the motor system and words relating to colors recruit visual areas, some studies also demonstrate that within motor areas neural activation levels are higher when participants are instructed to focus on the action associated with an object word than when they are instructed to focus on the object's color (van Dam, van Dijk, Bekkering, et al. 2012). Similarly, based on neuroimaging evidence, Hoenig, Sim, Bochev, et al. (2008, 1809) concluded that processing of a particular concept "does not selectively and constantly activate a specific sensory or motor region," since neural activation is based on contextual constraints, specifically the way in which the concept is cued, whether by using visual attributes or action attributes. Moreover, they used event-related potentials (ERPs) from electroencephalogram (EEG) experiments to rule out the possibility that this pattern of activation was a result of top-down modulation of conceptual processing, since the effects in question were present within 200 milliseconds of target onset (cf. van Dam, van Dijk, Bekkering, et al. 2012).[3] Collectively, such studies have led many researchers to conclude that neural activation in conceptual tasks is highly context-dependent and varies not just from person to person but also from occasion to occasion, depending on various contextual factors. This puts pressure on the attempt

[3] Both Hoenig, Sim, Bochev, et al. (2008) and van Dam, van Dijk, Bekkering, et al. (2012) use EEG to measure conceptual processing and claim that context-dependent effects pertain to conceptual processing proper since they occur within 200 milliseconds of stimulus onset. McCaffrey and Machery (2012, 273) dispute that researchers have a principled way of knowing that such effects are instances of conceptual processing even if they occur in such a short period of time after stimulus onset. As will emerge in due course, I think that such problems are just symptoms of the inherent difficulty involved in attempting to identify concepts with patterns of neural activations.

https://doi.org/10.1017/9781009223645.003 Published online by Cambridge University Press

to identify concepts with determinate neural correlates that are relatively constant in each individual, let alone across individuals.

This recent research builds on an older body of work that relies on behavioral evidence, which finds that participants' categorizations often vary with respect to context and are out of step with their own considered judgments, sometimes without warrant. In many classic studies on categorization, researchers found that the features associated with a given concept vary widely across experimental contexts. As we might expect, different features are associated with the same concept in different contexts (e.g. with the concept PIANO in the context of producing music and moving furniture; see Barclay, Bransford, Franks, et al. 1974), and many associated features in property verification tasks appear and disappear with context (e.g. the feature "has lungs" is associated with the concept BEAR in the context of the sentence, "The bear caught pneumonia," but not in other contexts; see Barsalou 1982). More drastically, psychologists have found that people's considered judgments about concept membership do not always accord with their categorizations of various instances on particular occasions. Among numerous other results, participants in some experimental conditions judge some odd numbers to be "more odd" than others and rank them on a scale of "oddness," even though they also agree in other conditions that the concept ODD NUMBER is all-or-none (Gleitman, Armstrong, & Gleitman 1983; cf. Barsalou 1983). Thus, just as the pattern of neural activations associated with a concept can vary considerably across contexts, the constituent features associated by an individual with a particular concept are also not constant in the different contexts in which that concept is accessed.

These experimental results suggest that there is good reason to doubt that a pattern of neural activation that is found to be associated with a concept in a certain experimental task is *constitutive* of that very concept, as opposed to merely being associated with it in some way. The variability in the way that concepts are manifested in different contexts cautions against equating neural activation on a given occasion with the manifestation of the relevant concept. However, those researchers who would identify the manifestation of a concept with some determinate pattern of neural activation, particularly in sensorimotor areas, could respond by saying either that: (a) there might yet be a conceptual "core" activation pattern common to all contexts; or (b) the concept might be identified with the *totality* of activations associated with the concept; or (c) there may exist some higher-dimensional invariance despite the apparent variability, and we might be able to extract some kind of constancy amidst the flux in the neural activation data. Each of these possibilities must be considered separately. When

https://doi.org/10.1017/9781009223645.003 Published online by Cambridge University Press

it comes to (a), some of the neuroimaging evidence already cited strongly indicates that there is no core common to the processing of many concepts. After surveying a body of research that relies on both behavioral and neuroimaging evidence, Lebois, Wilson-Mendenhall, and Barsalou (2015, 1772) argue for a "no-core" theory of concepts, concluding: "Rather than concepts containing cores that are activated automatically independent of context, concepts only appear to contain information that is dynamic and context-dependent." Similarly, Hoenig, Sim, Bochev, et al. (2008, 1801) write that there is a need to posit a strong form of conceptual flexibility "in order to account for the high degree of heterogeneity in category-related imaging findings with no single study detecting all the hitherto reported category-related activation foci …, and some studies even failing to find any category-related effects …" Some of this evidence counts not just against modal theories, but amodal ones as well. As Lebois, Wilson-Mendenhall, and Barsalou (2015, 1773) state: "the evidence we have reviewed suggests that concepts have no cores at all, amodal or modal." As for (b), identifying the concept proper with the totality of all such neural activations, this is problematic especially if there is no core to conceptual representations, since these activations will then be disjoint. A set of disjoint neural activations is a poor candidate for the neural correlate of a conceptual representation, at least if the aim is to effect a reduction of concept deployment and possession in neural terms. To put it in familiar philosophical terms, this would mean that concepts are multiply realized by their neural correlates. Finally, when it comes to (c), the possibility of discovering some higher-dimensional commonality in the variable patterns of activation, this is at present just a promissory note. It depends on the discovery of some non-obvious higher-order function that would serve to unite the disjoint neural correlates. Moreover, it would not seem to account for the variability in behavioral and verbal responses, which also points in the direction of fluidity and instability.[4]

Increasingly, many of the researchers involved in generating and interpreting the relevant data about the neural correlates of concepts insist that the correct interpretation of these results is the radical one that there is no stable, self-contained, context-independent neural representation that corresponds to any given concept. As Yee and Thompson-Schill (2016, 1016; original emphasis) put it: "… *the concepts themselves are inextricably*

[4] As will become clearer in the course of this chapter, I think it is a mistake to expect a single theory to explain fluidity and instability at the neural and behavioral levels, but researchers in this area often aim at a reductive account.

https://doi.org/10.1017/9781009223645.003 Published online by Cambridge University Press

linked to the contexts in which they appear, so much so that the dividing line between a concept and a context may be impossible to clearly make out …" These researchers also maintain that one should think of "the concept itself as changing slightly each time it is retrieved, and that there is no real demarcation between what is activated in a given instance and the concept itself" (2016, 1018). Connell and Lynott (2014, 393) perhaps state it most radically and most pithily: "one cannot, in effect, ever represent the same concept twice." Thus, a large body of recent research leads to the conclusion that the neural correlates of concept activation are highly variable and contextually determined.

2.3 Empirical Accounts: Cognitive Psychology

In cognitive psychology, two main theories of concepts have been dominant in the past few decades: (i) prototype theory, which holds that the content of a concept is based on a weighted set of features or stored exemplars, and (ii) theory theory, which posits that the content of a concept is based on its links to other concepts or its place in an encompassing theory. In this section, I will attempt to outline both theories and briefly introduce some of the empirical evidence that supports each. I will also try to relate these theories to those encountered in the previous section. Then, I will consider two ways of reconciling the two theories, and give some reasons to favor one over the other, which will allow me to introduce a different approach to concepts in Section 2.4. But before presenting each of these theories, I will briefly review the "classical" or "definitional" view of concepts that these two more recent views are usually regarded as displacing.

On a classical view of concepts, a concept consists in a definition, which supplies necessary and sufficient conditions for falling under the concept in question. To possess a concept is to have the definition in mind and to be able to deploy it in cognitive tasks such as categorization and inference. Moreover, to have the concept in mind or to activate it on a given occasion is to entertain the definition, and acquiring the concept consists in coming to have the definition. This view of concepts has been largely abandoned by philosophers and (especially) psychologists, for a variety of reasons, which are worth recounting very briefly. First, to say that concepts consist in and are grounded by their definitions seems to pass the buck, since those definitions presumably themselves consist in concepts (e.g. an apple is a fruit that is round, sweet when ripe, etc.). The natural way to stop this regress is to say that it bottoms out in a set of conceptual primitives, often

https://doi.org/10.1017/9781009223645.003 Published online by Cambridge University Press

thought to consist in a set of percepts and logical connectives. The percepts are then grounded in their connections to the sources of the relevant perceptual stimuli. But it is generally agreed that the empiricist philosophical project of building up concepts from percepts has largely failed and that our conceptual edifice cannot be constructed in this systematic fashion from percepts (cf. Fodor 1981). Another problem is that very few if any concepts are amenable to strict definitions. Any set of necessary and sufficient conditions that are put forward for defining even seemingly innocent and straightforward concepts are bedevilled by exceptions, and in numerous cases no amount of tinkering manages to avoid the problem (see e.g. Fodor 1981, 283–292). Third, even concepts that seem amenable to definitions at a particular stage of inquiry do not always retain those definitions as the inquiry progresses. This occurs time and again in intellectual history and the history of science, even though the concepts in question appear to persist, which suggests that definitions do not ground or provide identity conditions for concepts. Although the problem might be partly addressed by equating concepts with their definitions as determined *at the end of inquiry*, this would force us to admit that almost all our concepts are discarded and replaced with each definitional adjustment over the course of intellectual history. The same would hold for individual cognitive development, since the definitions associated with our concepts are often modified in the course of ontogeny, which would give rise to widespread conceptual replacement.[5] A fourth problem pertains specifically to concept possession: We are often warranted in attributing possession of a concept to a thinker even when that thinker is unaware of the purported definition, whether implicitly or explicitly (e.g. a schoolchild may have the concept CIRCLE despite not knowing that a circle is a plane figure all of whose points are equidistant from a given point). A fifth problem relates mainly to conceptual activation: Some empirical work indicates that entertaining a concept does not generally involve entertaining the concepts that occur in its definition (see e.g. Fodor, Garrett, Walker, et al. 1980). Moreover, other empirical results show that concepts are structured not in terms of definitions but in terms of features that are neither singly necessary nor jointly sufficient for concept membership, as we will see shortly in recapitulating some of the evidence in favor of the prototype theory. These are some of the main problems that have led philosophers and cognitive scientists to

[5] Some cognitive psychologists are willing to accept such a consequence (e.g. Gopnik 1984; 1988; Carey 1985; 1988; 2009), but I have argued elsewhere that they need not and ought not (see Khalidi 1998).

https://doi.org/10.1017/9781009223645.003 Published online by Cambridge University Press

conclude that concepts should not be equated with definitions and have compelled them to look for other theories of concepts.

The central claim of prototype theory is that concepts are structured not as definitions but as sets of representations of weighted features (see e.g. Smith & Medin 1981, 61–101). For example, the concept APPLE would be a list of feature representations (e.g. ROUND, EDIBLE, SWEET), each feature being weighted according to its importance to the concept. The more features an instance shares with the concept and the more important those features, the more typical that instance is of the relevant concept. Individual thinkers recognize and categorize objects as apples based on their having attained a certain requisite "score" that takes into account the weighting of each feature, not on the basis of having a set of necessary and sufficient features. More typical exemplars have more features, or more of the important features, and would attain a higher score than less typical exemplars.

The experimental evidence that led to the development of the prototype theory relies on behavioral rather than neural measures, including explicit judgments in categorization tasks, accuracy of responses in recognition tasks, and reaction times in both categorization and recognition tasks. The theory proceeds from the familiar idea that when it comes to many categories or concepts, some instances are generally considered more central, representative, or typical. For instance, at least among many North Americans[6], an apple is considered a more typical fruit than a coconut, a robin is regarded as a more typical bird than an ostrich, and a sofa is taken to be a more typical piece of furniture than an ottoman. This familiar fact is taken not just to be a matter of conventional wisdom but to reflect the way in which concepts are represented in the minds of individual thinkers. There are several different types of experimental findings that have been deemed to be evidence for typicality effects. One type of experimental result finds that for many concepts or categories, some instances are considered more typical by most participants when they are explicitly asked to rate them on a typicality scale. For example, participants are asked to rate the extent to which an instance represents their "idea or image of the category" (Rosch 1975, 199; cf. Rosch & Mervis 1975, 588). They are given words such as "orange," "apple," "pineapple," and "coconut," and are asked to rate their typicality as instances of the concept FRUIT on a scale from 1 to 7. Unsurprisingly, "orange" is rated more typical than "pineapple," which

[6] It is fairly uncontroversial that specific typicality results are culturally variable, but there is also some debate as to whether typicality results are found at all in some cultures; for discussion see Kemmerer (2019, 67–68) and references therein.

is rated more typical than "coconut," with a high degree of agreement among participants (undergraduates living in California) (Rosch 1975, 198). The results of this experiment demonstrate that some instances of a concept are generally regarded as more typical than others, but so far, this does not necessarily reveal anything about the structure of conceptual representations, since it may just reflect conventional or common knowledge about typicality. A second type of experimental result aims to establish that the instances of a concept that participants consider more typical are also those that are found to share more features with other instances of that concept. In this experiment, participants are given words such as "orange," "apple," "pineapple," and "coconut," along with the instruction to list as many features of that fruit as possible (e.g. "sweet," "juicy," "round," and so on) within 90 seconds. The fruits judged more typical (e.g. orange) are found to share more features with other fruits than the ones judged less typical (e.g. coconut) (Rosch & Mervis 1975, 582). A third type of result finds that participants are faster when categorizing those instances that are regarded as more typical. For example, participants are asked to respond "true" or "false" to such statements as "An apple is a fruit," "A pineapple is a fruit," and "A coconut is a fruit," and their reaction times and error rates are measured. Shorter reaction times are recorded for sentences involving the instances judged more typical (Rosch 1978, 38).[7] A fourth experimental result finds that superordinate concepts are better primes for typical instances than nontypical instances. When participants are asked to judge quickly whether two simultaneously presented words or images are identical, the word for the superordinate concept acts as a prime only for word pairs naming typical instances (Rosch 1975). As the experimenters put it: "Apparently, hearing the category name leads subjects to expect typical category items and not to expect atypical items" (Rosch, Simpson, & Miller 1976, 498). Finally, prototype theorists claim that the words for typical instances are likely to be named first and more frequently when subjects are asked to list instances of a certain concept, and the words for typical instances are the first ones to be learned by children and are learned more quickly by them (Rosch 1978, 38–39). Such findings are meant to support the prototype theory in various ways. For example, a concept instance that has more features or more heavily weighted features

[7] In the same series of experiments, participants were instructed that "some reds are redder than others" and that a Pekinese is a "less doggy dog" than a Retriever or a German Shepherd. The instructions also read, in part: "Don't worry about *why* you feel that something is or isn't a good example of the category … Just mark it the way you see it" (Rosch & Mervis 1975, 589; emphasis added).

https://doi.org/10.1017/9781009223645.003 Published online by Cambridge University Press

will be categorized more rapidly due to the fact that it achieves the requisite "score" in a shorter amount of time. Thus, typicality judgments are thought to reflect something about conceptual structure and the nature of conceptual representation.

The main competitor of the prototype theory is the theory theory of concepts, which has not been modeled as precisely as the prototype theory, but has been put forward as an alternative picture of conceptual structure.[8] According to the theory theory of concepts, concepts are embedded in a larger framework of explanatory beliefs (or theories), which thinkers draw upon in performing a particular cognitive task. Different parts of the entire corpus of beliefs may be deployed in different tasks, even ones involving a single concept. Proponents of the theory theory posit an interrelated network of conceptual information rather than self-contained collections of feature lists. Since the content of a concept depends in systematic ways on the contents of other concepts in a conceptual repertoire, this would make the theory theory a holist rather than an atomist account of concepts. As Murphy and Medin (1985, 289) once put it in a seminal article, referring to the then-dominant prototype theory:

> [C]urrent ideas, maxims, and theories concerning the structure of concepts … are inadequate, in part, because they fail to represent intra- and inter-concept relations and more general world knowledge. We propose a different approach in which attention is focused on people's theories about the world …

They go on to say: "we wish to reduce the importance of individual attributes [i.e. features] in conceptual representations and to emphasize the interaction of concepts in theory-like mental structures" (Murphy & Medin 1985, 292). This alternative picture is based on experimental evidence indicating that in many categorization tasks, thinkers rely not (or not just) on typical features or features that are statistically or probabilistically associated with concepts. Rather, they draw on theoretical, explanatory, and causal information related to the concept. Especially in non-routine categorization and inference, experimental participants forego the former type of information in favor of the latter. Much of the evidence for this theory first emerged from work in developmental psychology. Keil and collaborators carried out a series of experiments with children (kindergartners, second graders, and fourth graders) to investigate whether their concepts

[8] Later, in Section 2.4, I will argue that it is a mistake to view the theory theory as an alternative theory of conceptual structure and to view it as a direct rival to the prototype theory.

https://doi.org/10.1017/9781009223645.003 Published online by Cambridge University Press

can be identified with lists of features (as in prototype theory), or with theories (as in the theory theory). In these experiments, the typical setup consists in reading a short story to the children and then asking them questions concerning categorization. Experimenters sometimes repeat the story, ask the children to repeat the entire story, and encourage children to deliberate at some length and justify their categorization decisions. I will outline two well-known examples that illustrate the types of experimental protocol used. One story used by Keil (1989, 162) in his experiments described animals living on a farm who neigh, eat oats and hay, and are saddled and ridden by people. The animals are examined by scientists, who find that they have the insides of cows and the blood and bones of cows. Moreover, their parents and offspring are cows. Then children were asked what they thought the animals were, horses or cows. They were encouraged to justify their categorizations in conversation with the experimenters. Another story described taking a raccoon, shaving some of its fur, dyeing it black with a white stripe down its back, then inserting a sac of smelly odor into its body. Again, children were asked what they thought the resultant animal was, a raccoon or skunk. In the first case, most younger children said the animals were horses, while most older children said they were cows, and in the second case, most younger children said it was a skunk, while older children said it was a raccoon (Keil 1989, 164–182). The results of these experiments suggested to Keil and colleagues that for many concepts children exhibit a shift from relying on superficial features to relying on explanatory theories. This shift occurs as early as the preschool years for some concepts and as late as fourth grade for others. After the shift, and into adulthood, thinkers use causal theories in performing categorization tasks rather than relying on characteristic or typical features (Keil 1986; 1989).

Evidence for the theory is not restricted to developmental studies. Other empirical work demonstrates that concept learning in adults occurs more readily for concepts that have features that are correlated by causal connections rather than ones that are not correlated in this way. For example, adults show a strong tendency to cluster features among which a causal link can readily be made rather than ones that do not exhibit this tendency. In learning hypothetical disease categories in a concept-learning experiment, participants found it easier to link dizziness to earaches and weight gain to high blood pressure, rather than dizziness to weight gain and earaches to high blood pressure (Murphy & Medin 1985, 302). Other evidence comes from work on conceptual combination: Combining nouns denoting concepts to form a compound concept often requires thinkers to draw on background beliefs rather than simply select overlapping features. For example, an *expert*

https://doi.org/10.1017/9781009223645.003 Published online by Cambridge University Press

repair is a repair done by an expert, whereas an *engine repair* is (probably) not a repair done by an engine but a repair done to an engine (Murphy & Medin 1985, 306). We understand such conceptual combinations because we have theories about experts, engines, and what is involved in carrying out repairs.[9] Finally, there is considerable evidence showing that in feature-listing experiments, the features that people choose to list in connection with a concept vary widely with context in conformity with people's background theories. As mentioned in Section 2.3, different features are associated with the concept PIANO in different contexts. Similarly, when it comes to the concept NEWSPAPER, if one specifies the context of building a fire, the feature "flammable" may be listed, but not in other contexts (Barsalou 1993). To generate such features, people draw on broader theoretical information. Thus, the theory theory considers concepts to be embedded in theories and implicated in informational structures with considerable causal content, and this account is at loggerheads with the one provided by prototype theory, which conceives of concepts as clusters of features.

In this section, we have encountered two currently dominant theories of concepts in cognitive psychology, and in the previous section, we outlined two competing theories of concepts in cognitive neuroscience. There might appear to be a natural alliance between these two opposing theories and the two theories we encountered in the previous section. In particular, prototype theory seems to be compatible with modal or sensorimotor accounts of concepts, since the features associated with many prototype concepts are often posited to be perceptual representations.[10] If the features are not modal, some theorists speculate that these features are themselves concepts that can in turn be represented as sets of features, and that this process will eventually terminate in purely sensorimotor features. Of course, the claim that all concepts can ultimately be decomposed into percepts is a long-standing tenet of empiricism, which is why some philosophers have referred to this research program in cognitive science as "empiricism" (Weiskopf 2007) or "neo-empiricism" (Machery 2007;

[9] There is a large body of work on conceptual combination that I will not be able to summarize or address here, but it should be mentioned that some of it purports to show that the prototype theory accounts for the phenomena better than the theory theory (see e.g. Hampton 1997; Hampton 2006).

[10] See Barsalou (2016, 1133) on this point: "Compressed representations, such as CCRs [cross-modal conjunctive representations], are essentially the same kind of representations as prototypes in cognitive theories of concepts … According to prototype theory, statistically likely features are extracted from category exemplars and conjoined in a prototype that represents the category conceptually. Notably, prototypes are not amodal symbols arbitrarily linked to exemplars. Instead, the features of exemplars appear in the prototype that covers exemplars, following various possible forms of data compression, as for CCRs."

https://doi.org/10.1017/9781009223645.003 Published online by Cambridge University Press

McCaffrey & Machery 2012). If this program could be carried out, it might provide a way of linking the prototype theory to modal theories of concepts. The empiricist program to construct concepts out of percepts is widely thought to have failed, as mentioned earlier in this section, but even if it could be made to succeed, much more would have to be done to show how the non-perceptual aspects of prototype theory (e.g. feature weights) could be implemented neurally without introducing amodal elements and undermining the main claim of modal theories. Meanwhile, the theory theory seems consistent with accounts of concepts that emphasize their context-dependence and flexible structure, since researchers who consider concepts to be embedded in theories tend to think that different portions of those theories may be active in different contexts, as seen above. But relating these theories more directly is a daunting task, since it is one thing to say that both theories emphasize flexibility and context-dependence but quite another to show how one might map onto the other. Moreover, both theories face the same challenge of somehow discerning some fixity amidst the instability. Hence, whatever resonance exists between modal theories and prototype theories (on the one hand) and flexible theories and theory theories (on the other) is merely suggestive and the connections between them are, at least for the time being, somewhat tenuous.

In providing an account of the structure of concepts, these theories also effectively supply individuation or identity conditions for concepts, or ways of distinguishing one concept from another, or of saying what makes something the concept that it is. For example, on a prototype theory of concepts, what makes something a concept of APPLE is that it consists in representations of the features associated with apples, each weighted based on its importance or centrality to the concept. Moreover, what it is for the concept to be accessed or deployed on a given occasion is for those representations to be jointly active (along with their accompanying weights) in some way. Similarly, on the theory theory, the concept APPLE is often identified with the theory in which it inheres. That is what makes it the concept that it is, and the concept is accessed when that theory (or some significant part of it) is activated. These brief hints already suggest that both theories face considerable challenges in providing individuation or identity conditions for concepts. In particular, identifying concepts across individuals or even within individuals across times is not a trivial task in either case. For example, on the prototype theory, one would have to say which or how many of the features have to be held in common, and how similar the weightings would have to be, for different structures to be representations of the same concept. The difficulties are at least as formidable

https://doi.org/10.1017/9781009223645.003 Published online by Cambridge University Press

on the theory view: Does every change in theory lead to a change in concepts, and if not, how much change or difference can one tolerate? But setting aside these difficulties for the time being, I want to argue that these theories may not be rival theories of concepts at all, but might be tracking different kinds.

The debate between the prototype theory and the theory theory has led to something of a standoff in cognitive psychology. In response to the impasse, various philosophers and cognitive scientists have argued that the prototype theory and theory theory are not genuine rivals. Following Weiskopf (2009, 168), we can divide these theoretical proposals into two groups: pluralist theories and hybrid theories. The basic difference is that pluralist theories posit *different kinds* of conceptual representation, while hybrid theories consider that there is just *one kind* of conceptual representation incorporating distinct components that include different types of information. Hybrid theories typically claim that the component of a conceptual representation accessed on any given occasion depends on the task and that each concept consists of a single complex entity consisting of two or more different representations in different representational formats. Sometimes hybrid theories are supposed to include rule-based and similarity-based representations rather than theory-based and prototype-based representations (e.g. Close, Hahn, Hodgetts, et al. 2010). Pluralist theories, on the other hand, claim that different theories about concepts in cognitive science are actually discussing different kinds of entities, and that each of our concepts may be associated with two or more different representational types. I will now put forward two related considerations for supporting a pluralist account, one based on the methods used by the two theories and the other based on their respective explananda. In earlier work (Khalidi 1995), I argued that the prototype theory and the theory theory rest on different bodies of evidence involving disparate experimental methods. When it comes to the prototype theory, much of the experimental evidence derives from tasks that require rapid categorization of stimuli or snap judgments made under time constraints without a surrounding context. By contrast, results that support the theory theory mainly derive from experimental setups that involve explaining and justifying classifications or inferences in the context of a broader cognitive exercise (e.g. listening to a narrative) and they do not usually include time constraints or measures of reaction times. Accordingly, I hypothesized that the prototype theory and theory theory were pitched at different "levels of explanation" (bearing in mind the caveats from Chapter 1 against conceiving of levels hierarchically). With the benefit of over a quarter century of additional

https://doi.org/10.1017/9781009223645.003 Published online by Cambridge University Press

empirical work, this hypothesis still seems plausible, since the results that are taken to support the two theories remain largely disjoint and the types of behavioral and verbal measures are of a different order. Neither theory has been displaced by the other and there has been no real convergence between them. A second source of support for the claim that the prototype theory and theory theory are not genuine rivals derives from the fact that they are focused on different explananda, namely concept *activation* and concept *possession*, respectively. Even though the prototype theory has the resources to provide an account of concept possession, as I briefly indicated above, the methods used are supposed to gauge activation on a particular occasion. By contrast, the theory theory is primarily focused on assessing concept possession. Much of this empirical work compares children with adults or children at different developmental stages. The main objective is not to assess conceptual activation at a particular moment, but to issue a verdict about concept possession by certain individuals or groups based on the totality of evidence gathered. Hence, given that the methods used are largely disjoint and the explananda are distinct, I would maintain that these theories are not addressing the same phenomena.

It may be objected here that the theories are more plausibly interpreted to be addressing different explananda regarding the *same kinds*, rather than different kinds altogether. Just as we might have a theory that explains what it is for a creature to possess a heart and another theory that explains what a heart does, the theory theory explains what it is to possess a concept and the prototype theory explains what it is to activate a concept. But the analogy is not apt, since the two theories of concepts do not even agree on how to individuate their subject matter. It is not as though prototype theory accepts the theory theory's account of concept individuation and possession, and then proceeds to tell us how those very entities are activated on different occasions. Moreover, the causal processes being investigated in each case tend to be somewhat different. The prototype theory is best at explaining categorization and recognition over short time scales when words or images are shown to participants in the absence of a broader theoretical or practical context. The theory theory tends to be better at accounting for reflective or deliberative categorization and inferential judgments that require integrating information from various sources, often involving longitudinal studies.[11] This adds some credence to the hypothesis that the

[11] These differences might also be related to "dual-systems" or "dual-process" models of cognition, which posit two cognitive systems, a fast, implicit, and automatic system and a slower, explicit, and more deliberate system. Prototype theory accords with the type of rapid thought processes

https://doi.org/10.1017/9781009223645.003 Published online by Cambridge University Press

prototype theory and theory theory are not rival theories of concepts, but are tracking different kinds. The causal processes that they engage in may interact or intersect in various ways, as I will try to indicate in the next two sections, but they remain somewhat distinct. Moreover, I will argue in the following section that some versions of the theory theory individuate concepts partly on the basis of etiology, or with reference to their causal antecedents, and not just with reference to synchronic causal powers. This lends further support to the conclusion that they are investigating different cognitive kinds. This means that, strictly speaking, there are no such things as concepts simpliciter; for clarity, it may be best to give at least one type of entity a different label.

2.4 A Functional Account of Concepts

Philosophical discussions of concepts (or word meanings) have been dominated in the past several decades by the divide between internalism (or individualism) and externalism, which was outlined in Section 2.1. As already mentioned, externalists hold that concepts are individuated by determinants that are external to the mind of the agent, while internalists maintain that the subject's internal perspective is what individuates a concept. When it comes to the concept APPLE, roughly speaking, the externalist says that what makes it the concept that it is are the thinker's causal connections to the apples in the thinker's environment, whereas the internalist holds that the concept is individuated by its intrinsic character (e.g. the elements of the definition, or the features in the prototype, or the tenets in the theory). One problem with the way these positions are often characterized is that it is not sufficiently emphasized that what matters to the externalist is not so much what is currently in the thinker's vicinity but rather what was causally efficacious at the time of concept acquisition. The various thought experiments proposed to motivate the externalist position usually identify the concept with an external determinant in the context of acquisition, not the context in which the concept is accessed or activated, even though this is not always clearly articulated. Thus, externalists effectively individuate concepts *etiologically*, based on their ontogenetic causal history.

associated with the fast and automatic cognitive system (System 1), whereas responses associated with the theory theory are more in keeping with the slower and deliberate cognitive system (System 2). On some accounts, these two cognitive systems operate independently and can issue in different responses to certain cognitive challenges such as categorization or inference (see e.g. Evans & Stanovich 2013). Moreover, these systems are often regarded as belonging to different kinds or types of cognitive system, not just different token systems (Samuels 2009b).

https://doi.org/10.1017/9781009223645.003 Published online by Cambridge University Press

While many cognitive psychologists are avowedly internalists about concepts, most philosophers profess externalism.[12] The reason that psychologists (and neuroscientists) tend not to embrace externalism is not difficult to ascertain. Externalism distinguishes among concepts (e.g. APPLE, ORANGE, etc.) based on their causal origin or history, whereas most scientific disciplines and subdisciplines seem to categorize phenomena on the basis of their causal powers or efficacy. More to the point, if one's aim is to explain behavior, there would appear to be a commonality to behaviors exhibited by individuals who share internal or "narrow" states. This commonality is of interest to psychology when it comes to understanding, say, the discriminatory abilities of individuals or the distinctions they make, without regard to the underlying reality or social context (see e.g. Block 1986). It has therefore been difficult for philosophers to convince some psychologists to take externalism seriously when it comes to the individuation of concepts, at least explicitly (for a seminal exchange illustrating the depth of the divide, see Rey 1983; Smith, Medin, & Rey 1984; Rey 1985). Notwithstanding the fact that most philosophers are externalists, some have been swayed by reflecting on scientific taxonomy or psychological explanation to say that internalism or individualism is the only sound taxonomic strategy in science (e.g. Fodor 1987). But this may be because they have been misled by reflecting on some scientific disciplines to the exclusion of others. It is true that in large swathes of the natural sciences, what matters for taxonomic purposes is not the diachronic features of the phenomena being investigated but rather their synchronic properties, specifically their causal powers. Chemists do not usually distinguish among molecules of glucose based on whether they have been artificially synthesized in the lab or are the result of photosynthesis in plants. However, in many other sciences, including some of the physical sciences, some taxonomic categories do track causal origin, trajectory, or history. I have argued elsewhere (Khalidi 2021) that etiological kinds are widespread in science, notably in such sciences as cosmology, geology, and biology, for a number of bona fide scientific reasons (cf. Burge 1986, 18–19). If we grant that there is nothing in principle to prevent scientific inquiries from individuating phenomena based on etiology, does this apply to cognitive

[12] Hampton (2006, 84) writes: "Although many philosophers … have identified major difficulties with descriptivism, preferring to fix conceptual contents in terms of extensions (an Externalist theory of concept individuation), the large majority of cognitive psychologists still subscribe to this basic descriptivist position." According to the PhilPapers 2009 Survey of Philosophers: 51.1 percent favor externalism about mental content, 20.0 percent internalism, and 28.9 percent other.

https://doi.org/10.1017/9781009223645.003 Published online by Cambridge University Press

science in particular? And if so, why should a scientific inquiry into concepts need to individuate them based on their causal origin or history?

Over the past few decades, philosophers have made a strong case for the claim that folk psychology, at least in some circumstances and for certain purposes, individuates concepts based (at least partly) on etiology. But that does not mean that cognitive science does. Is there any evidence that scientific research programs do so, and moreover, that they have good reason for doing so?[13] Some work in developmental psychology mentioned in the previous section, in which psychologists attempt to ascertain whether or not children possess concepts of certain animals or artifacts, are arguably continuous with our folk psychological practices of concept attribution. The same goes for many other research programs in cognitive psychology, social psychology, and educational psychology. In many such domains, there is a concerted effort to determine which concepts individual thinkers possess or have mastered in various experimental conditions. When developmental psychologists conclude that a kindergartner has acquired the concepts ALIVE and DEAD (Bascandziev, Tardiff, Zaitchik, et al. 2018; cf. Carey 1985), or a preschool child possesses the concepts ANIMAL and ARTIFACT (Greif, Kemler Nelson, Keil, et al. 2006), or a two-year-old has the concept CAUSE (Waismeyer, Meltzoff, & Gopnik 2015), or a three-month-old infant has the concepts CAUSE, COST, and GOAL (Liu, Brooks, & Spelke 2019), they are doing so, in large part, on the basis of certain synchronic causal abilities that they have, that is to say, their responses, discriminations, preferences, and related behavioral and cognitive capacities. But they are also basing themselves, at least partly, on etiology. That this is the case can be seen by looking a little more closely at some of the examples just cited.

When a three-month-old infant is ascribed the concepts CAUSE, COST, and GOAL, this is done on the basis of rather minimal discriminatory capacities, having to do with looking times at certain experimental stimuli. By comparing the durations of looking times at different scenes showing goal-directed behavior, experimenters conclude that infants at such an early age possess the concepts in question. Briefly, infants generally exhibit longer looking times at inefficient compared to efficient (or cost-effective) goal-directed causal behavior on the part of other human agents, indicating that they are surprised by such inefficient behavior. This in turn, signals to the researchers that they are able to make the distinction

[13] Burge has been perhaps the most vocal exponent of the view that a science of psychology is and ought to be externalist (or anti-individualist, to use his term). This has been a consistent theme in his work from Burge (1986) to Burge (2010), but he focuses mainly on perception and perceptual states.

https://doi.org/10.1017/9781009223645.003 Published online by Cambridge University Press

between behaviors that are cost-effective and those that are costly. Yet, the researchers admit that their experiments and those of other developmental psychologists do not reveal "how richly … infants represent the costs and goals of other people's actions" (Liu, Brooks, & Spelke 2019, 5). In other words, the infants are able to make some of the relevant discriminations though they might not possess complex representations. Work on children at twenty-four months shows that infants are able to make causal inferences, which implies that they are not only distinguishing instances of causation from non-instances, but deploying the concept CAUSE to infer a cause–effect relationship between two events, indicating further progress along the conceptual trajectory (Waismeyer, Meltzoff, & Gopnik 2015).[14] These two-year-olds still do not have a sophisticated understanding of causation but based on their experiments, the researchers claim that they can distinguish causation from correlation in intervening on the world. In a similar vein, Greif, Kemler Nelson, Keil, et al. (2006) conclude that preschool children make a distinction between animals and artifacts and have the corresponding concepts (viz. ANIMAL, ARTIFACT), based on the fact that they ask different types of questions when confronted with unfamiliar exemplars from each category. In their experiment, three- to five-year-olds posed more questions and made more guesses about functions and behaviors for artifacts than animals, whereas they made more category guesses and asked more questions about niche or location for animals than artifacts. These researchers ascribe the concepts of ANIMAL and ARTIFACT to preschool children while acknowledging that they "may not be able to verbalize the abstract differences between causal patterns associated with living kinds and with artifacts" (Greif, Kemler Nelson, Keil, et al. 2006, 459). Finally, a body of research in developmental psychology shows that while preschoolers have an undifferentiated concept ANIMATE/ACTIVE/REAL, which maps onto the word "alive," older children undergo a process of conceptual change, after which they acquire the concept ALIVE. In addition, Bascandziev, Tardiff, Zaitchik, et al. (2018) demonstrate that some six-year-olds can be induced to acquire the concepts ALIVE and DEAD with training. Even though a thorough mastery of the concepts may not occur ordinarily until much later, they show how certain kinds of training can result in acquiring these concepts for some six-year-olds. In all three cases discussed, while psychologists ascribe concepts to children based partly

[14] This work does not make explicit mention of the concept CAUSE, though it is clear from the experimental results that the children are engaging in causal inference, which (at least) involves possessing the concepts CAUSE and EFFECT.

on their causal powers of discrimination, recognition, categorization, and inference, at the same time, the concepts ascribed go beyond their bare abilities when they are "narrowly" conceived. One might well ask why the experimenters attribute richer concepts to the children than may seem warranted by their narrow responses and behavior.

There would appear to be at least two reasons that cognitive psychologists judge that children possess concepts like CAUSE, COST, GOAL, ANIMAL, ARTIFACT, ALIVE, and DEAD, rather than a suite of more rudimentary concepts that reflect their (presumably) more impoverished understanding, recognitional capacities, discriminatory abilities, and so on. The first reason is that children and adults inhabit a single world and children are in contact with the same external stimuli as adults. Possession of a concept marks a certain cognitive achievement and indicates that the thinker in question is able to successfully interact with and navigate some aspect of the world, and psychological theories aim to explain these abilities. Hence, concepts can be seen to be anchored in their origins in our shared world. It is true that psychologists also aim to understand the differences in the ways that different human thinkers conceive of the world (children and adults, members of different cultures, speakers of different languages), but these differences can be described against a background of shared concepts. Indeed, as many philosophers have pointed out, these differences can *only* be described if we presuppose a base of shared concepts (e.g. Davidson 1974). Another reason is that the children are on a developmental trajectory that will, in the overwhelming majority of cases, result in their becoming competent users of language and attaining the full-blown adult versions of these concepts. Rather than ascribe different concepts at every developmental stage, and attribute an entirely different conceptual repertoire at each stage, psychologists regularly say that the children possess these concepts, yet they have an incomplete understanding of them. It is always possible to say something like: The child has the concept of ANIMAL, she just does not know very much about animals. Since these children are in the process of mastering these concepts and becoming full-blown members of our linguistic community, we use the resources of natural language to capture their thoughts, and this practice involves using the concepts lexicalized in the language of their community. Thus, despite rare explicit assent to externalism by cognitive scientists, some research programs in cognitive psychology appear committed to individuating concepts not just on internalist or individualist grounds, but also externalistically, and they have good reasons for doing so. The rationale for this resides partly in the fact that these thinkers inhabit the same environment from which these

https://doi.org/10.1017/9781009223645.003 Published online by Cambridge University Press

concepts derive and to which they apply, as well as the fact that they are or will become part of the same linguistic community.

If concepts in cognitive science are ascribed to agents based partly on synchronic causal powers and partly on etiology, both factors playing a role in determining whether an individual thinker possesses the relevant concept, that also indicates a certain continuity with folk psychological practices of concept attribution. Even though externalists tend to emphasize causal origin as fixing the content of concepts and thoughts, our ordinary folk psychology does not ignore internalist features altogether. According to a pure or strict externalist account of concepts, what makes my concept APPLE that very concept is its etiology, in other words, my causal history with actual apples and with other language users. On a standard externalist account, if I have acquired the word "apple" by way of direct or indirect contact with apples and with members of my linguistic community, I thereby possess the concept APPLE and my word "apple" means APPLE. But if I am under severe misapprehensions about apples and think that they are brown, starchy, and grow underground, and cannot discriminate them from potatoes, then it would be unusual in most everyday circumstances to ascribe to me the concept APPLE. In ordinary conceptual ascriptions, we do indeed give considerable weight to etiology when it comes to the individuation of concepts, but not to the exclusion of synchronic factors, namely the behavioral and cognitive causal powers of individual thinkers. Folk psychology gives some weight to the synchronic abilities of agents, particularly their recognitional, discriminatory, and inferential capacities. Sometimes externalist and internalist accounts are thought to deliver two different concepts of concept, call them concept$_E$ and concept$_I$, or two notions of conceptual content, wide and narrow content. But there is another way to conceive of the situation: Concepts are possessed by individuals in virtue of etiology *and* causal power, and concepts are individuated by both factors in tandem. This goes as much for folk psychology as for many areas of cognitive psychology, as I have tried to argue. Admittedly, I have used just a few examples from the voluminous empirical literature to corroborate this claim, but the methods used and the reasoning deployed are very widespread. Moreover, I have deliberately chosen relatively recent examples from influential research groups to indicate that this is representative of current practices of concept attribution in cognitive psychology, which builds on a body of work on concepts and conceptual development undertaken over the past several decades.

Having argued that much work in cognitive psychology can be understood to individuate concepts both internalistically and externalistically,

https://doi.org/10.1017/9781009223645.003 Published online by Cambridge University Press

I want to propose some convergence between this approach and at least some versions of the theory theory of concepts, which was discussed in the previous section. If the theory theory is understood not as a theory of conceptual structure, but rather as a theory of concept possession, then the account of concepts that is implicit in the research programs cited above can be reconciled with it. One main claim of the theory theory is, as Carey puts it, that concepts "must be identified by the role they play in theories" (1985, 198) and "conceptual role at least partly determines the content of concepts" (2009, 502). This functional or causal-role approach to individuating concepts and determining concept possession can be reconciled with externalism, at least if functions are identified not just narrowly but widely.[15] That is, when ascertaining whether a thinker possesses a certain concept, one attends not just to their categorizations narrowly construed but also to the relevant environmental causes and etiology. By contrast, proponents of the prototype theory, who are more interested in questions of concept activation, tend to endorse an internalist account of concepts (e.g. Hampton 2006). Since etiological individuation need not coincide with individuation according to intrinsic causal powers, this further supports the claim that the two approaches are identifying different cognitive kinds, as I argued in the previous section. Moreover, as I have already suggested, these two theories can be seen to be pitched at different levels of explanation, as these levels were characterized in Chapter 1: relatively "closed systems," each of which is causally integrated and somewhat autonomous from other systems. One way to characterize the difference between the two theories is in terms of what Marr (1982, 22–31) labels the "algorithmic" and "computational" levels, respectively.[16] Marr's theoretical framework can be used to provide an additional reason for considering prototype theory to be pitched at a different level from theory theory. That is because it is natural to think of prototype theory as providing an algorithm for concept activation. As mentioned in the previous section, prototype theory provides a procedure for activating a concept once the weighted values of the associated features have reached a certain threshold.

[15] To be more precise, I am arguing for a hybrid (narrow-wide) construal of functions. A hybrid theory of functions of this kind has been articulated by Griffiths (1993), who applies it to the case of biological functions as well as human artifactual functions.

[16] In much earlier work (Khalidi 1995), I hypothesized that the "levels" in question correspond to what Dennett has called the "design stance" (prototype theory) and "intentional stance" (theory theory). I now think that Marr's framework provides a better basis for understanding the differences. Kitcher (1998) has suggested that there is a natural convergence between Marr's algorithmic level and Dennett's design stance. She also identifies a convergence between Dennett and Marr

https://doi.org/10.1017/9781009223645.003 Published online by Cambridge University Press

By contrast, the theory theory of concepts is more closely related to Marr's "computational level." At the computational level of analysis and explanation, the emphasis is on what concepts enable cognitive agents to achieve and why they possess the concepts that they do. For instance, infants who acquire the concepts CAUSE, COST, and GOAL are able to make certain discriminations and discover certain things that they would not be able to do without that concept. As the researchers put it: "An early-emerging sensitivity to the causal powers of agents, when they engage in costly, goal-directed actions, may provide one important foundation for the rich causal and social learning that characterizes our species" (Liu, Brooks, & Spelke 2019, 17747). Moreover, infants possess these concepts, at least in part, because they inhabit a world of causal agents who are attempting to achieve their goals while incurring low costs, or because their ancestors did so and were naturally selected to think in these terms. A number of philosophers have interpreted Marr's computational level as having an externalist dimension (see e.g. Burge 1986; Kitcher 1998; Egan 2014; Shagrir & Bechtel 2017). In particular, Kitcher (1998, 14) claims that Marr's project shows that science need not be methodologically solipsist, since his computational level "makes essential reference to factors beyond the subject's skin in characterizing psychological states." Similarly, Shagrir and Bechtel (2017, 209) write that the "why aspect" of the computational level "forces researchers to look to the structures in the world that the organism engages through its visual system." Not only are environmental and etiological factors causally relevant in explaining cognitive processes at this level, it also bears repeating that they enter into the *individuation* of cognitive kinds. If we are interested in reliability and success in navigating and interacting with the world, then we will need to individuate concepts partly in terms of their environmental causes and etiology. Moreover, at the computational level, the aim is not to give a structural account of concepts (as the prototype theory does), but to give a functional account, in terms of what possessing a concept enables a thinker to do.

I have been arguing that concepts are cognitive kinds individuated in terms both of synchronic causal powers and etiology. Concepts endow their possessors with certain causal powers or cognitive abilities, though it would be hopeless to try to assign a specific set of canonical abilities

when it comes to the implementational level and physical stance, respectively. However, Kitcher (1998, 14–15) proposes that Dennett's "intentional stance" is a fourth level, distinct from Marr's "computational level," and she criticizes Dennett for holding that the intentional stance cannot be rendered scientific, since Marr's program shows that it can be.

https://doi.org/10.1017/9781009223645.003 Published online by Cambridge University Press

to each concept. Moreover, each concept operates only in conjunction with other concepts, so there is no requisite set of abilities attached to each concept in isolation from others. In addition, I have argued that concepts are also identified with their contextual and historical determinants and individuated partly in terms of those determinants. How are these two determinants of concepts, synchronic and diachronic, related to each other? Rather than think of these two factors as distinct components of conceptual content, they can be conceived as joint determinants of conceptual content, combining to provide individuating conditions for concepts. This gives us a way of transcending a dilemma faced by "two-factor" or "dual-factor" theories of concepts (e.g. Block 1987; Carey 2009), at least if these factors are understood as distinct components of concepts.[17] Since the internalist factor and the externalist factor tend to pull in opposite directions whenever a thinker exhibits significant ignorance or harbors a misconception, that is, whenever one's thoughts are seriously out of step with the world or the linguistic community, it is not clear what dual-factor theories would say about a thinker's concept in such cases. Is it the narrow concept or the wide concept that takes precedence? Rather than thinking of concepts as being resolvable into two possibly opposing factors or components, it is more in keeping with both folk and scientific taxonomy to regard concepts as amalgamating both factors. But if concepts are individuated both synchronically and diachronically, do we have specific criteria for how much disparity in causal powers to tolerate, or how much leeway to give those whose concepts have the right etiology, in individuating concepts or determining which concepts are possessed by a thinker? There seem to be no hard-and-fast criteria, whether in folk or scientific psychology, but that should not undermine the entire enterprise of conceptual taxonomy. In many cases in science, such as in assigning species to higher phylogenetic taxa, we weigh diachronic and synchronic factors in making a determination about classification, without having sure-fire rules for doing so. Here too, we balance different considerations and issue in a verdict as to whether the thinker has the concept in question or not. As in other cases in scientific taxonomy, there may be borderline cases in which there is no fact of the matter as to whether a thinker possesses a concept or not. Sometimes

[17] Block sometimes talks in terms of narrow and wide "determinants" of meaning rather than aspects or kinds of meaning (see e.g. Block 1986, 620), but it is not clear how he views the relationship between them. The account I am proposing may be closer to the "long-armed" conceptual role semantics of Harman (1982).

https://doi.org/10.1017/9781009223645.003 Published online by Cambridge University Press

the etiology is right but the synchronic causal powers are lacking; at other times, the agent is making the appropriate discriminations and categorizations but the external antecedents are not the ones we expect.[18]

Before concluding this functional account of concepts, it is necessary to say something about the role of language in individuating concepts. In finding the right words to capture a thinker's thoughts, we are not simply forcing a fluid mental reality to conform to a rigid representational medium. The process of allocating the correct linguistic labels to a thinker's thoughts is not like trying to measure temperature using a crude and inaccurate thermometer, because language is a representational medium that allows us to capture subtle differences among thinkers whose concepts might not seem to be in perfect alignment. Earlier, I said that cognitive psychologists decide to attribute rich adult-like concepts to infants and children rather than ascribing a suite of more impoverished concepts. Developmental psychologists choose not to report the children's beliefs in terms of some alternative set of concepts, CAUSE*, GOAL*, ANIMAL*, ARTIFACT*, and so on, but instead rely on the familiar set of adult concepts. That is not a distortion of their mental lives because language enables us to pinpoint areas of disagreement using a common stock of concepts. So when describing the mental lives of preschool children, instead of saying that they have some concept ANIMAL* rather than ANIMAL, we can say that they have the concept ANIMAL but they do not know that animals are multicellular, or that they have a common lineage, or that they breathe oxygen, or whatever other discrepancies we might find between our views and theirs. The same goes for any two thinkers who harbor different conceptions, make distinct associations, or hold disparate beliefs in connection with any given concept. In some cases, we do resort to introducing novel concepts and coining new terms when faced with thinkers whose behavior and responses cannot be captured at all in concepts expressed in natural language, but these are the rare exception, even when it comes to infants and other atypical thinkers. (One such case was mentioned above: Carey attributes an undifferentiated ANIMATE/ACTIVE/REAL concept to preschoolers.) Language provides a way of fixing the individuation conditions of concepts despite variability in the way they are deployed across contexts

[18] I do not have space here to address the notorious cases that have figured so prominently in the philosophical literature, such as the WATER or ARTHRITIS cases, which are used to motivate externalism. But in at least some of these cases, there may be no determinate answer to the question of concept possession. Recall that on this hybrid account, it is not enough simply to be causally related to a certain environment or community, one also needs to have certain powers of discrimination, inference, and so on.

https://doi.org/10.1017/9781009223645.003 Published online by Cambridge University Press

or by different individuals. By giving a role to the words of shared, public, natural languages in the individuation of concepts, one incorporates the *social* etiology of concepts as well as their *natural* etiology or their causal origin in a shared world.

Some cognitive scientists claim that words are what give the *illusion* of conceptual fixity or relative stability (e.g. Casasanto & Lupyan 2015), but this is misguided not only because we can accurately capture subtle individual differences using language as just indicated, but also because language can be seen to have a role in the development of concepts. As Lupyan and Thompson-Schill (2012, 20) put it: "Words may matter far more for conceptual representations than previously considered, in that some concepts may only attain sufficient 'coherence' when activated by verbal means." Language can be seen to play an active role in the evolution of conceptual thought, both ontogenetically and (more speculatively) phylogenetically. When it comes to ontogeny, a large body of research shows that words support individuation, inductive inference, and causal reasoning. For example, providing twelve-month-old infants with a word highlights commonalities among objects that go undetected in the absence of a word (see Waxman & Gelman 2010 and references therein). This same research demonstrates that the word-concept link is not just one of pure association and that words do not merely play the role of an attentional spotlight. In fact, "the conceptual status of words comes not from the sound of a word itself, but rather from its role within the linguistic and social system in which it is embedded" (Waxman & Gelman 2010, 107). As for phylogeny, the evidence is more speculative, but there is also reason to think that language played a prominent role in the development of full-fledged human conceptual thought. Many philosophers have theorized about the indispensability of language for augmenting cognition and enabling certain conceptual achievements that would not have been possible without language. Clark (1998, 173–174) writes:

> The role of public language and text in human cognition is not limited to the preservation and communication of ideas. Instead, these external resources make available concepts, strategies, and learning trajectories which are simply not available to individual, un-augmented brains.

On such a view, the evolution of language and the evolution of the human conceptual repertoire are intertwined in such a way that it is difficult to prise them apart. Our concepts coevolved with the words of natural languages, and language provides part of the scaffolding that enables human conceptual thought. If linguistic symbols play an active role in shaping mental phenomena, both ontogenetically and phylogenetically, then it

https://doi.org/10.1017/9781009223645.003 Published online by Cambridge University Press

is misleading to think of language as providing the *illusion* of stability to concepts. Rather, linguistic symbols are instrumental in determining conceptual identity. This provides another reason for casting doubt on the notion that language is an inadequate representational medium that distorts the rich content of thought.

To sum up, I have argued that in both folk and (at least some areas of) scientific psychology, concepts are individuated with regard to both synchronic causal powers and diachronic etiology. This method of individuating concepts pertains primarily to investigations of concept possession (as opposed to activation), which are pitched at what Marr designated the computational level of explanation. Moreover, this approach is most closely aligned with the theory theory of concepts, by contrast with the prototype theory, which is pitched at the algorithmic level. Since these different theories pertain to different explanatory levels, I proposed that they investigate different scientific domains, which are populated by different kinds. Rather than a unitary cognitive kind, *concept*, these theories are attempting to understand different constructs, call them *concept*$_1$ and *concept*$_2$. A question that arises here is how to relate the hypothesis that there may be different kinds associated with the label "concept" in these different domains to recent evidence from neuroscience, particularly the theories surveyed in Section 2.2. I argued in Section 2.3 that despite the fact that there might seem to be a close alliance between the theories proposed on the basis of neural evidence and those based primarily on behavioral measures, the connections are likely more tenuous. In fact, the neural evidence pertains to what Marr dubs the "implementational level." The modal theory of concepts is clearly framed in implementational terms since it posits that the very same neural networks that process perceptual information also process concepts and that they do so using the resources of the sensorimotor systems in the brain. Rather than expect a direct correspondence between this neural or implementational theory and theories framed in algorithmic (or computational) terms, modal theories of concepts can be seen to operate with different taxonomic categories. What they are calling "concepts" might not correspond either to the construct identified at the algorithmic level or that investigated at the computational level, but some third construct, *concept*$_3$. Hence, I am proposing that the theories of concepts under discussion in this chapter are investigating (at least) three different kinds that are implicated in distinct causal processes.[19] This may seem like a profligate agenda to multiply concepts beyond necessity, but the

[19] Machery (2005; 2009) also argues that there is a plurality of concepts of CONCEPT in contemporary cognitive science, but he takes this as grounds for eliminativism about concepts and for denying

https://doi.org/10.1017/9781009223645.003 Published online by Cambridge University Press

alternative is to suppose that there is a neat correspondence among the entities tracked by these different research programs with their different methodologies and their investigations into disparate causal processes, some of which are individuated with reference to context and history while others are not, and some of which aim at explaining reflective thought and inference while others are primarily interested in recognition and spontaneous judgment. It would be more surprising to discover that they were all really theorizing about the very same kind of thing.

2.5 Objections and Replies

In the previous section, I tried to outline an account of concepts that considers them to be real kinds in cognitive science, in accordance with the account of cognitive kinds that I elaborated in Chapter 1. I can anticipate a number of objections that the account may elicit, so I will try in this section to raise and respond to the most prominent ones. Since many of the replies are already implicit in the previous discussion, I will address them rather briefly and without detailed elaboration.

One objection would draw attention to the similarities between this theory of concepts and the theory theory of concepts. Given that the theory theory is prone to some well-known objections, it is natural to ask whether this account is open to the same objections. Two of the most powerful objections to the theory theory are the circularity objection and the holism objection. The circularity objection says that the theory theory equates concepts with theories or parts of theories, but theories are themselves constituted at least in part by concepts. This leads to a "mereological paradox" (Laurence & Margolis 1999, 44). The response to this objection is to say that concepts should not be *equated* with theories or parts of theories. Concepts are inextricable from theories and are not independent of them, but that does not mean that they are identical to miniature theories or sets of theoretical tenets. If concepts are individuated in part narrowly, based on a thinker's ability to perform certain cognitive tasks associated

that *concept* is a natural kind. I will not try to respond directly to his arguments, but my main disagreement with Machery is that I think that there is a single kind that plays a dominant role in higher cognitive processes such as those described in this section, while he thinks that three different alleged kinds play this role (in addition to prototypes and theories, he discusses exemplars). Machery (2009, 242–243) thinks that each of these categories corresponds to a kind, but withholds the term "concept" from any one of them. By contrast, I have argued that there is a continuity between our pre-theoretic notion and the category that features in some scientific research programs, which justifies reserving the term for that category.

https://doi.org/10.1017/9781009223645.003 Published online by Cambridge University Press

with those concepts, then a possessor of a concept holds certain theoretical tenets associated with it, whether explicitly or implicitly. But that does not make the concept equivalent to those theoretical tenets. This account of concepts is a functional rather than a structural one; part of the mistake is to conceive of the theory theory as being a structural rival to the prototype theory. The holism objection is related; it says that if a concept is associated with a theory or part of a theory, then any change in theory results in a change in the concept, with the result that concepts differ constantly within and across individuals. I would respond to this worry by briefly expanding on what I said in the previous section regarding concept individuation. Cognitive scientists (as well as the folk) usually represent a thinker's mental life using the resources of natural language. The terms of natural language align the thinker's concepts with those of the community and any differences between them can be represented against a background of common concepts. Hence, concepts do not differ incrementally with every difference in belief, and concepts do not change slightly with every change in belief. Differences and variations in thought can be captured by indicating disagreement among beliefs using a common stock of concepts.

Another objection to this theory of concepts might question its affinity to a rather different account, namely the interpretivist account associated with Dennett's "intentional stance" (e.g. Dennett 1987). That account of concepts is widely perceived to make concept ascription subjective or response-dependent, which many cognitive scientists doubt is compatible with a true science of the mind. In previous work (Khalidi 1995; Khalidi 1998), I favored an interpretivist account of concepts, which is indeed response-dependent. But even though the interpretivist account contains important insights (including the inextricability of concept and theory, or meaning and belief), I now think that response-dependence does not accord with scientific or folk taxonomic practice. I have argued that concept possession is partly a matter of a thinker's synchronic causal abilities (sorting, recognizing, discriminating, categorizing, inferring, and so on) and partly a matter of the thinker's relations to external determinants (including social determinants) in the history of the thinker. These determinants of concept individuation do not depend ultimately on how others assess them, but on facts about thinkers themselves and their relations to the world.

A different objection to a broadly functional account of concepts would question whether one can in fact equate the content of each concept with its functional role. The reply to this concern is that this is not meant to be a reductive account. The individuation of concepts is based on what individuals do with those concepts, taking etiology and environment into

https://doi.org/10.1017/9781009223645.003 Published online by Cambridge University Press

consideration. But there is no attempt to identify a requisite set of abilities corresponding to each concept. There is no fixed set of hoops that a thinker must jump through to demonstrate that he or she possesses the concept APPLE or ORANGE (or for that matter, CAUSE, ANIMAL, or LIFE). There may be different ways of exhibiting possession of a concept, and anyway, most concepts are acquired in close confederation with other concepts. It is not a coincidence that many studies in cognitive psychology examine a small number of closely related concepts together (e.g. CAUSE, GOAL, and COST; CAUSE and EFFECT; ANIMAL and ARTIFACT; ALIVE and DEAD). Hence, there is no question of isolating each concept and specifying its functional role in isolation from other concepts. This does not mean that cognitive scientists do not have ways of gauging whether individual thinkers possess certain concepts as opposed to others, as already suggested in looking at some of the empirical work on children's concepts. But as is clear from considering this work, concepts should be thought as abilities or *capacities* rather than *objects* or concrete particulars. In Chapter 1, I said I would indicate, where applicable, how cognitive kinds fit into certain overarching superordinate kinds. In this case, I am calling for revising a common ontological view among some philosophers and cognitive scientists that takes concepts to be objects instead of capacities.[20] As I have already argued, these cognitive capacities should be individuated functionally, where functions are understood in hybrid terms, both causally and etiologically.

It may also be objected that the account of concepts that I have outlined is too abstract to be suitable for cognitive science. In particular, it does not seem to square with the usual cognitive scientific claim that concepts are to be identified with *mental representations*. I think the right response to this objection is to say that concepts can be abstracta and mental representations at once. They are abstract kinds, but they are also representational in nature, since they are individuated by what they represent. This representational content is supplied jointly by their causal history and their causal role, though I have not tried to supply a full-blown theory of mental representation. However, concepts should not be thought of as concrete entities or objects with a definite location or spatial dimensions. That is what we would expect to find at the implementational level of explanation, not at the computational level. In some discussions in cognitive science "mental representation" is used synonymously with "neural representation," but the question of the existence and nature of neural representations is an

[20] This view is perhaps most clearly instantiated by Fodor (1998, 2–4), who regards it as a fundamental error on the part of cognitive science to think of concepts as abilities rather than objects.

https://doi.org/10.1017/9781009223645.003 Published online by Cambridge University Press

independent one, and moreover, I have given reasons for doubting that there will be a neat correspondence between the computational and implementational levels.[21]

There is a related objection that would cast doubt on the claim that the construct "concept" identified in these different research programs really picks out different kinds. It may seem both taxonomically extravagant and prima facie improbable that completely different kinds have (unwittingly) been identified by these different research programs.[22] To this objection, I would reply that there may well be important relations between kinds identified at the implementational, algorithmic, and computational levels (though I cannot be sure that real kinds have already been identified in this vicinity at the first two levels). But we should not expect a relation of reduction or a one-to-one correspondence between them. To make this more plausible, consider the concept ALIVE, as discussed at the computational level. As we have seen, current research in cognitive psychology concludes that this concept is connected with such concepts as DEAD, is typically acquired by children by ages ten to twleve (in North America), can be induced via training to younger children, enables children to answer questions and make distinctions concerning animals and plants, and is associated with a "vitalist" theory in biology (Bascandziev, Tarfdiff, Zaitchik, et al. 2018). Meanwhile, other researchers in psychology have noted that children and adults respond distinctively to stimuli that move in certain ways, reacting to them as though they are alive (see e.g. Mandler 1992). This automatic response to (apparently) living things is clearly not tantamount to judging that something is truly alive nor that the concept ALIVE or LIFE correctly applies to it. It would seem to pertain to a system or mechanism that takes as input certain perceptual stimuli and issues a fast and non-deliberative response, perhaps something at the algorithmic level. Moreover, the relationship between the output of that system and the considered categorization judgment is likely to be highly indirect. We often react immediately to a stimulus (a branch shaking in the wind, a piece of fluff propelled by a draft) as though it were alive, but go on to suppress our initial reaction. This response may be one input into the conceptual system but that does not mean that it is a component of

[21] For a recent account of neural representation, see Shea (2018).

[22] While advocating a pluralist view of concepts, Weiskopf (2009) makes the case that the different kinds are subordinate kinds belonging to a single superordinate kind, *concept*. I do not have space to counter his arguments for this conclusion directly, but I think the considerations that I have brought forth argue for the existence of entirely different kinds. See Machery (2009, 243–245) for some replies to Weiskopf's arguments.

https://doi.org/10.1017/9781009223645.003 Published online by Cambridge University Press

the concept itself. In the specific context of the representation of animacy, Mandler proposes a number of different levels from the perceptual to the conceptual and hypothesizes that "human infants represent information from an early age at more than one level of description" (1992, 602). There is no reason to conclude that the representations at each level simply correspond to or are combinations of those at lower levels.

A final objection would accuse this account of concepts of being maximalist or intellectualist rather than minimalist, to revert back to a distinction from Section 2.1. Given the tight link that I have posited between concepts and language, this theory of concepts would seem to apply exclusively to language-using creatures, or at least those creatures who are on track to acquire language. While some philosophers would be sympathetic to such a view, many others would not, claiming that there are no good grounds for denying concepts and conceptual thought to, say, nonhuman animals (for detailed discussion, see Camp 2009). In the previous section, I argued that concepts are ontogenetically and phylogenetically associated with language. Language is not just a representational medium that happens to individuate concepts, it is coeval with them. Therefore, even though non-linguistic thinkers can possess concepts, it is doubtful that their concepts can be individuated as finely as those of language-using creatures, and there will be far more inherent vagueness in the concepts that they possess. This does not mean that non-linguistic thinkers cannot have concepts, just that there will be more cases in which it is genuinely indeterminate which concepts they possess.

2.6 Conclusion

The causal-etiological account of concepts that I have outlined in this chapter suggests that concepts are a real kind in our cognitive ontology. To briefly recap the main conclusions of this chapter, it may be worth reframing them in terms of the central research questions that were outlined in Section 2.1. When it comes to the first question about the identity of concepts, I have argued that in the cognitive or computational domain, concepts should be considered abstract functional entities rather than concrete particulars like neural structures or processes. Since they have intentional content, they can be thought of as mental representations as long as they are not considered concrete representational entities. As for the individuation of concepts, their representational contents derive from their wide functions and are a combination of etiological and causal factors. This means that concept acquisition is not just a matter of the thinker having the right causal history but also

https://doi.org/10.1017/9781009223645.003 Published online by Cambridge University Press

being able to make the right distinctions and inferences – though there is no set of requisite abilities or achievements associated with each concept. This, in turn, implies an account of concept possession according to which having a concept is a matter of a thinker's inferential and discriminatory capacities, as well as their etiology. Finally, this view of concepts does not address the question of concept activation, which I have argued is more properly addressed at the implementational level.

Another way of framing this account of concepts is in terms of some of the central theoretical divides mentioned in Section 2.1. I have defended a hybrid account of conceptual content that combines internalist and externalist elements. I have also argued against the modal (or sensorimotor) theory of concepts, at least as a theory of concepts at the computational level, rather than at the implementational or algorithmic levels. This account has a strong affinity with the theory theory of concepts, though I have also argued that there may be a different type of psychological entity at the algorithmic level that may conform to the prototype theory. Like the theory theory, the account is holistic about conceptual content. Some standard objections against holism have been briefly addressed. I have argued against a response-dependent account of concept identification and individuation, on the grounds that concept possession is determined by a thinker's abilities and history, not ultimately by the responses of other members of their communities. Finally, the account lies between minimalism and maximalism, in that language may not be strictly necessary for the possession of concepts, but it helps determine the fixity of concepts as well as their acquisition and possession.[23]

It may be thought that this chapter has focused on kinds of *individual* concepts, not on the kind *concept* as such. Never mind the cognitive kinds APPLE, ORANGE, CAUSE, ANIMAL, and LIFE, what about the cognitive kind *concept* itself? What are its identity conditions and how does one individuate it? An analogy can be drawn here with the kind *species* and its subkinds, *Panthera tigris* (tiger), *Drosophila melanogaster* (fruit fly), and *Acer palmatum* (Japanese maple). Note that in both cases, these subkinds do not bear a straightforward relationship of subordinate kinds to a superordinate kind. In the case of species, the higher taxa, like genera and families are true superordinate kinds. By contrast, the kind *species* and the kind *concept*, bear a relation to their subkinds (particular species and particular

[23] Camp (2009) distinguishes three grades of conceptual thought: minimalist, moderate, and intellectualist. The moderate conception seems to be the most congenial to the account that I have defended in this chapter.

https://doi.org/10.1017/9781009223645.003 Published online by Cambridge University Press

concepts, respectively) that is more akin to the determinate–determinable relationship, like the relation of *color* to *violet*, or the relation of *length* to *one millimeter*. Thus, if we grant that specific concepts (e.g. APPLE, ORANGE) are causal-etiological cognitive kinds, then there is no further question as to whether *concept* itself is a real kind in the cognitive domain. If the determinates are real kinds so is the determinable. Still, if specific concepts are identified with their functional roles, what are concepts *in general*? I have argued that cognitive scientists have ways of individuating concepts such as CAUSE and ANIMAL and means of assessing whether a thinker possesses them, but do they have ways of discovering whether a thinker has concepts at all, as opposed to not having concepts? The synchronic and diachronic conditions for possessing concepts are closely related to the conditions for having beliefs or thoughts. I will not attempt to identify the conditions for thought or cognition or general, which are notoriously difficult to spell out. But there is some consensus that cognition involves an ability to represent the world in a stimulus-independent way (see e.g. Camp 2009; Beck 2018). For full-fledged *conceptual* cognition, mere stimulus independence may not be enough. Camp (2009, 303–304) argues that conceptual thought combines stimulus-independent representational ability with the capacity to recombine representations, and that this "makes a practical difference for achieving the most basic aim of thought: using information about the world to solve problems and facilitate one's survival and flourishing." Hence, conceptual thought can be distinguished from thought in general by this ability to represent the world in a stimulus-independent way and recombine those representations in such a way as to interact flexibly with the world. This characterization also combines synchronic and diachronic factors, since it involves the ability to navigate the world flexibly as well as the ability to form representations that derive from the world. It also confirms the importance of language for concept possession, since language is a medium that facilitates representation and keeps track of representational recombination, even though language may not be strictly necessary for having concepts. As I argued at the end of the previous section, there is nothing to preclude all thinkers, including nonlinguistic creatures, from having concepts. But in the case of nonlinguistic thinkers, it will often be difficult to determine whether they possess concepts, and there may be no determinate answer to the question of which concepts they possess.

https://doi.org/10.1017/9781009223645.003 Published online by Cambridge University Press