1. Introduction
This paper explores a theory of concept acquisition that aims to appease tensions in the debate between Concept Empiricism and Concept Nativism. The debate focuses on lexical concepts –ideas of kinds and properties that tend to be named with single words in Natural Languages. Examples include ideas of things as being apples, tigers, water, cars, pizza, knowledge, chopping, and quarks. Presently the discussion is limited to lexical concepts of natural kinds, and only those of objects and substances whose instances are perceptible. According to the theory I offer, the Baptism of Mental Terms (BMT), many such concepts can be acquired by forming a mental description that appeals to patterns in perceptual experience and then assigning a new simple mental name to the property that is picked out by the mental description. The theory is a result of considerations from the philosophical and psychological debate on concepts, combined with theories of natural kind terms from the philosophy of language.

Many theories of concepts adopt a Representational Theory of Mind (RTM), and we will adopt that framework as well for the purpose of this paper. Usually, theories under the RTM assume what I call the Building-Blocks model, on which concepts come in only two classes –Primitive representations and Composite representations. In this paper, I challenge that assumption by offering a counter-example. I show how to make room for a third class of concepts, resolving key tensions between Concept Empiricism and Concept Nativism.

The BMT is closely connected to the Kripke/Putnam/Burge/Soames process for assigning meanings to linguistic terms. The general idea isn't completely novel; many discussions of the linguistic process gesture at a mental version, either as a direct mental analog of the linguistic version, or else as a prerequisite part of the linguistic process. It is only by developing a detailed model, however, that we see just what is involved cognitively in carrying out such baptisms. Working out a model, moreover, is what reveals this overlooked solution to the on-going concepts debate.

2. The Building-Blocks Assumption
According to the RTM, thoughts are structured representations, analogous to linguistic representations, with the simplest units of meaning composing together (in accordance with some conditions).

---

1I do not presently discuss concepts of artifact kinds, social kinds, event kinds, state kinds, or theoretical kinds.
2What is on offer is not a metaphysical account of what it is to have a concept. Whatever it is to have a concept, the BMT process offers a pathway to that state.
3It is worth exploring how the observations made here apply to non-representational theories of concepts. These include Brandom (1994), Dennett (1977), Dummett (1993), Churchland (1981), Matthews (2007); and Peacocke (1992), taking concepts to be abilities, and Frege, (1892) taking them to be abstract objects.
4Kripke (1972), Putnam (1975), Burge (1979), Soames (2002).
'grammar', or composition rules) into more and more complex units of meaning. The Building-Blocks Assumption is a picture of how mental representations combine that furthers the analogy. On this picture, there are Primitive representations, which are the basic building blocks of thought, and there are Composite representations built up from the Primitives. The Composites inherit their meanings completely from their representational parts, and they do so by a rational-causal process, such that they are inferentially related to their parts. Both sides of the debate agree that full propositional thoughts (e.g., A BLUE APPLE IS NEXT TO A ROUND RED SNAIL) as well as phrasal concepts (e.g., BLUE APPLE, IS NEXT TO, and ROUND RED SNAIL) are among the Composites. I use the convention of capital letters for concepts.

The Primitives ground all of the meanings in some basic way. There are several different accounts of that basic way, but almost all of them appeal to a relationship between the representation and its cause. The cause is usually taken to be either a distal stimulus (something in the outside world) or else a proximal stimulus (retinal stimulation or sensations themselves). These accounts further spell out just what that causal relationship is, as in Dretske’s (1981) informational account, or Fodor’s (1990) Asymmetric Dependence account, or Millikan’s (1984) teleo-functional account. The important thing is, almost everyone under the RTM agrees that the Primitives are brute-causally related to their meanings—the world and our brains are innately built so that certain properties in the world cause these representations to trigger. They are the representations we get directly from perception (e.g., the taste of sourness or the experience of redness, if those are representational, or else whatever representations come as a direct result).

I find that the Building-Blocks picture forces a conflation of the acquisition and meaning-determination factors for representations. I pull these factors apart in my characterizations of Primitive and Composite.

<table>
<thead>
<tr>
<th>Composite Representations:</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Acquired) Acquired from more primitive representations through a rational/inferential process.</td>
</tr>
<tr>
<td>(Complex) Complex in representational structure, in the sense that their meanings are inherited from the meanings of the more primitive representations that were involved in their acquisition.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Primitive Representations:</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Innate) Possessed at birth, or else acquired only by brute-causal (i.e., non-rational, non-inferential) processes.</td>
</tr>
<tr>
<td>(Simple) Simple in representational structure, in the sense that they directly represent their meanings, rather than inheriting them from any other representations.</td>
</tr>
</tbody>
</table>
Note that even John Locke (1690) distinguished Primitive and Composite representations in this way – the sense in which he took perceptual representations to be ‘acquired’ is the same as the brute-causal sense of ‘innate’ that we are using here.\(^5\)

Within the Building-Blocks framework, certain questions in terms of the model arise. In particular, the debate between Concept Empiricism and Concept Nativism is almost completely about where to draw the line between the Primitives and the Composites, focusing on the lexical concepts. Recall that the lexical concepts are, roughly, the ones that tend to be expressed by single words, the smallest units of meaning in most natural languages. These include the concepts BLUE, APPLE, NEXT-TO, and SNAIL.\(^6\) Concept Empiricists usually put lexical concepts with the phrasal representations, claiming that they are Composite (acquired from other representations and built up from the representations involved in their acquisition). Concept Nativists put them on the side of perceptual representations, taking them to be Primitive. Empiricists believe, for example, that the concept APPLE is acquired and built from more simple representations, perhaps RED, ROUND, CRUNCHY, and SWEET, which in turn may be acquired and built from further Primitives. For Nativists, APPLE is among the sensory representations, set up at birth, or through brain development, to be triggered by apples in the world.

\[
\begin{align*}
\text{(Concept Empiricism)} & \quad \text{Most lexical concepts are Composite.} \\
\text{(Concept Nativism)} & \quad \text{Most lexical concepts are Primitive.}
\end{align*}
\]

There are plenty of good reasons to adopt the Building-Blocks framework. First is the observation that we only entertain a finite number of thoughts in our lifetimes, yet the number of possible thoughts that any human being can have is in(de)finite. The idea that we store some basic set of thought constituents and then compose the constituents together as needed explains this observation. Another thing we notice is that full propositional representations, like FURRY DOGS EAT DOG FOOD and WHITE CATS EAT CAT FOOD, seem clearly to be Complex, and Acquired by composing their more basic phrasal representations, like FURRY DOGS and CAT FOOD. The phrasal representations clearly break down also, into FURRY, DOG, CAT and FOOD. Since the Building-Blocks framework makes sense for the composition of propositional and phrasal representations, it seems natural to suppose that the same model of composition would be used when considering the lexical concepts.

The intuitions behind the Building-Blocks framework have been made more precise in the literature, and are known as the systematicity and productivity of thoughts.

---

\(^5\)Jerry Fodor’s (1981, 1985, 1998) notion of innate simples and John Locke’s (1690) notion of acquired simples both count as ‘Primitives’ in this sense. Their views diverge in ways that will become clear later – Fodor takes most lexical concepts to be among the Primitives while Locke takes them to be among the Composites.

\(^6\)Note that the notion of a ‘lexical concept’ involves no commitment to the agent’s having a natural language term that corresponds to it. Lexical concepts are simply the mental representations for properties that tend, in most languages, to be expressed by single words when and if expressed.
The Systematicity of Thoughts: Many mental representations have meaningful parts, and the parts contribute the same meaning to the complex representations in which they appear. For example, the concept DOG contributes the same meaning to the complex representations FURRY DOG and DOG FOOD.

The Productivity of Thoughts: From a finite set of concepts, we come to be able to have in(de)finitely many complex concepts. For example, from TREE, BIRD, and FOOD, we can have BIRD FOOD, BIRD TREE, TREE BIRD FOOD, and so on.

The Building-Blocks model is a natural way to account for these observations. Primitive representations compose together into Composites, building more and more complex meaningful representations.

One additional observation that supports the model comes from the Computational Theory of Mind (CTM) that is shared by many theorists under the RTM. On the CTM, an account of at least some of our thought processing appeals to the relationship between thought structures and our thought processing. Our deductive reasoning, most clearly, seems to be sensitive to the constituents of thoughts that the Building-Blocks model posits.⁷

The Logic of Thought Processing: Much of the cognitive processing that operates over thoughts is a lot like formal logic, which is sensitive to sentence parts. We infer, for example, from thoughts of the forms ‘P’ and ‘if P then Q’ to thoughts of the form ‘Q’, and from thoughts of the forms ‘All-x, if Px then Qx’ and ‘Pa’ to thoughts of the form ‘Qa’.

Even when we pull apart the acquisition and meaning-determination factors, it is hard to imagine how a representation could be any other way than Primitive and Composite as described here. In later sections I show that the Building-Blocks framework is to blame for much of the tension in the concepts debate, and that the factors can be pulled apart in at least one way that seems just as natural.

3. The Concepts Debate in Terms of Building-Blocks
In this section we reflect on the two pre-theoretic intuitions that give rise to the tension between Empiricism and Nativism. After that, I consider some further observations that, to me, clearly favor those intuitions over their counterparts. Again, the dispute is about lexical concepts –Nativists say they are Primitive and Empiricists say they are Composite. The lexical concepts are, roughly, the representations that correspond to words, like APPLE, WATER, SNAIL, RED. Again, I will only discuss

lexical concepts for natural kinds in this paper, and only natural kinds of objects and substances (not events or states).\(^8\)\(^9\)

Keeping with the analogy between thought and language, it seems natural to regard the word-sized representations, the lexical concepts, as Simple. We certainly feel as though we perceive and think about the world at that level. We see cars and people on the street, and hear trains and dogs, and our thoughts seem to be about things in terms of such categories. Human beings start out as Naïve Realists. That is, before reflecting on the matter, we feel as though we are in direct contact with the contents of our thoughts. Indeed, even expert philosophers and psychologists find it natural at least to take a Naïve Realist stance when ordinarily thinking about and experiencing the world. The tenet Simple, also known as Direct Reference theory, is as close to this Naïve Realist stance as one can get, while still accepting that our access to the world is mediated by representations. Historically this has always been the intuitive, default view. Giving in to the intuition that lexical concepts are Simple, Nativists take along the view that they are Innate.

The trouble is, the intuition that many of our lexical concepts are Acquired is equally strong and historically accepted as the intuition for Simple. Consider the concepts SNAIL, MUSHROOM, DINOSAUR, and COYOTE. There is a strong intuition that these concepts have been acquired as a result of our experiential contact with the world. It is almost impossible to imagine how we could have these concepts innately, or why we would. Moreover, it seems like the world gives us good reasons to think that things in the world fall into such categories. The concepts seem to be acquired not brute-causally, but by a rational process that takes patterns in experience as evidence for the existence of such kinds. We seem to learn through experience about the kinds and properties that populate our environment, that there are such properties as being an apple, being a cactus, being copper, and being a star. We also feel as though we can recall times at which we lacked some of these concepts, and sometimes we even seem to remember the moments at which we acquired them. Moreover, the reasoning is defeasible, and is often retracted in light of further evidence, as with WITCH and FLOGISTON. A little reflection adds the observation that human beings seem able to adapt to quite varied environments. It seems preposterous to suppose that we were born with representations directly for these properties. There are far too many, perhaps infinitely many, possible concepts that would count as lexical. It would be wasteful if not impossible, it seems, to have them all innately. This isn’t a point about evolution; it is simply a point about good design. Of course, on the Building-Blocks framework, if lexical concepts are acquired from patterns in sensory experience, they must have those sensory representations as part of their representational structure. This is a large part of the reasoning that brings Empiricists to believe that lexical concepts are Complex. But that forces them against the strong Simplicity intuition; Empiricists have to say that when we think that we are thinking directly about apples and mushrooms, we are really thinking about complex compositions of

\(^8\)I use ‘natural kind’ only to pick out the kinds that are typically thought to be natural, not in a technical or theoretical sense.
\(^9\)I do not even mean to suggest that there are such things. I do not presently discuss mental representations for particulars (like Plato and Arizona) nor for logical operators (like Or and If).
properties like red, round, crunchy..., and white, smooth, curved... This result is extremely counter-intuitive.

To resolve the tension, something has to give. Empiricists choose to hold onto the intuition for **Acquired** and abandon **Simple**. After all, they may rationalize, **Complex** is a very straight-forward account of how concepts can be **Acquired**. Nativists, on the other hand, want to preserve the intuition for **Simple**, so they accept as following from the logic that lexical concepts are **Innate**. In the debate, then, Nativists try to strengthen the intuition for **Simple** and weaken the intuition for **Acquired**. Empiricists try to strengthen **Acquired** and weaken **Simple**.

On the side of Empiricism, arguments are mostly aimed against **Simple**. Some arguments point to inference relations, such as seemingly analytic truths (All cats are animals, All apples are fruit) and inferential relationships between perception and concepts (Barsalou, 1999; Prinz, 2002). If the concept CAT is **Complex**, and built from the concepts ANIMAL, FURRY, and PURR, we would have a natural explanation for these inference relations.

Perhaps the most powerful argument against **Simple** comes from Frege's observation that some lexical terms may be causally connected to the same natural kind, yet carry different information. Consider the following variant on Frege’s puzzle about identity statements. Sammy is a young child who has two mental names that refer to coyotes. One of these mental names, F, was acquired through his visual perceptions of coyotes (suppose he saw some of these skinny-legged wolf-like animals at a zoo). The other mental name, G, was acquired through Sammy’s auditory perceptions of coyotes (suppose he heard some screeching howls through his window at night). If Sammy were to eventually learn that Fs and Gs were the same kind of animal, and form the thought ‘Fs are Gs’, he would presumably have learned something new. The thought ‘Fs are Gs’ is informative in a way that the thought ‘Fs are Fs’ is not. Since the external referents of F and G are the same, it seems like the only way to account for the difference in information is that the concepts are **Complex** and built up from different perceptual representations.

On the side of Nativism, observations have been made that seem to support **Innate**. One is that children seem to lock onto the right properties very quickly when explicitly being taught words for them. The perception of just one or two good examples of an apple, seems to be enough for children to lock onto the property of appleness. Such little experience doesn’t seem to be enough for learning to occur, certainly not by the formation and testing of hypotheses, so children must be innately wired to detect appleness. Another major argument for **Innate** is the one that comes from **Simple** and ‘the unlearnability argument’ (aka the ‘the standard argument’) given in Fodor (1975, 1981). The idea behind this argument is really the same as what I have called *Fodor’s Challenge*, the challenge of showing how a **Simple** concept can be **Acquired**.

---

10 Frege's Puzzle is introduced in his (1892).
In addition to defending Innate, problems have been raised against Complex. Problems with the definability of lexical concepts have been observed at least since Plato (see also Fodor, Garrett, Walker, and Parkes, 1980). Empiricists have tried to maintain Complex without definability, with Prototype Theory\footnote{Wittgenstein (1953, 1958), Rosch and Mervis (1975), Rosch (1978), Murphy (2002), Barsalou (1999), Prinz (2002).}, Exemplar Theory\footnote{Smith and Medin (1981).}, and Theory-Theory\footnote{Carey (1985), Gopnik and Meltzoff (1997), Keil (1989), Spelke (1994).}. But concepts that are Complex in these ways struggle to compose in ways that preserve the productivity and systematicity of thoughts (see Fodor, 1998). Against Complex, one could also point out the implausibility of many of the conceptual priorities that it requires. The Empiricist explanation for the seemingly analytic statements, like ‘All cats are animals’ is that ANIMAL is part of CAT. If this is the case, then children have to have the concept ANIMAL before they can have the concept CAT. However, the same explanation would have to hold for the analyticity of ‘All cats are mammals’. While it might seem plausible that children have ANIMAL before they have CAT, it seems absurd to think that children have MAMMAL before CAT. In similar vein, there doesn’t seem to be any particular set of representations that are required in order to possess any given concept. It seems a blind child may have the concept CAT without ever having seen a cat and a deaf child may have the concept without ever having heard a cat. Notice that we have no trouble thinking about the Cheshire Cat, the cat with only a mouth from Lewis Carroll’s fiction Alice’s Adventures in Wonderland. What makes something a cat, it seems, is simply that it is a cat.

As one can imagine, many moves are made, back and forth and back again, new observations are found, and there doesn’t seem to be any way to settle things conclusively. It is clear that the initial intuitions, the one for Simple and the one for Acquired, are both far too strong for either to vanish. I propose that both sides have too quickly become defensive in response to the tension. Neither intuition should be so easily dismissed. It is at least worth exploring the prospect of resolving the tension with an account on which many lexical concepts are Acquired and Simple. It will not do, however, merely to massage and adjust these notions into compatibility. Something will have to give, and the only thing that is obviously left is the Building-Blocks Framework. I propose that we challenge the assumption that a concept is Acquired if and only if it is not Simple. In particular, I would like to explore the possibility that Simple concepts can be Acquired.

4. The Baptism of Mental Terms: Abandoning the Building Blocks Model

In spite of its general endorsement within the concepts debate, there is no a priori reason to think that there are only these two classes of mental representation – Composite and Primitive representations. It is hard to imagine a representation not falling into these two categories. Moreover, and more centrally for this paper, it is hard to imagine how Complex and Acquired could come apart, especially as spelled out in Section 2. How can a rational process, i.e., a meaning-preserving process, bring about a concept that doesn't inherit those meanings? It is worth exploring alternatives, however. Resolving the tension between the simplicity intuition and the acquisition
intuition is what I am here calling Fodor's Challenge. This paper works out an account of concept acquisition that meets the challenge, offering one way for concepts to Acquired and Simple. I call it the Baptism of Mental Terms (BMT), as it is inspired by Kripke's (1972) baptism account for the coining of linguistic terms.

4.1 Sketch of the BMT Model
For reasons that will become clear shortly, I limit the scope of the theory as presented in this paper to lexical concepts for natural-kind properties. According to the Baptism of Mental Terms, most such concepts are neither Primitive nor Composite. Instead, they are the result of a process that is a mental version of the process in the Kripke/ Putnam/Burge/Soames account of the baptism of natural-kind terms in language. Lexicalized mental terms for natural kinds are acquired by forming a mental description in terms of representations already in possession to pick out a natural-kind property, and then baptizing a new simple mental symbol to be used as a representation for the property that is picked out by the description.

Baptized Mental Terms:
\[
\begin{align*}
\text{(Acquired)} & \quad \text{Acquired from more primitive representations through a rational/inferential process.} \\
\text{(Simple)} & \quad \text{Simple in representational structure, in the sense that they directly represent their meanings, rather than inheriting them from any other representations.}
\end{align*}
\]

The mental description is formulated in terms of representations already in possession by the agent. Once a new property is picked out by the description, a new simple mental name is baptized as the mental name for the property. This resulting mental name becomes a representation for that property.

In this minimal illustration, we imagine an agent that has a built-in perceptual system that takes inputs from the world and presents objects as having some color value between red, yellow, and blue, and some shape value between rectangular and round. In other words, we can say that the agent has a 2-dimensional perceptual space, with one dimension for color and one dimension for shape, and every perceived object falls somewhere in that perceptual space. The agent might then be said to have representations making up the two dimensions that are Innate and Simple, corresponding to these colors and shapes. See Figure 4.1.

---

14 Eric Margolis (1998) and Robert Rupert (2001) have offered models that aim to abandon the Building-Blocks assumption, but they come short of appeasing the tension between the Empiricist's intuition that concept acquisition involves a rational/inferential process and the Nativist's intuition that concepts are simple in representational structure. Margolis fails to show that the concepts acquired on his model are representations of the right properties. Indeed, he barely discusses the properties that lexical concepts are supposed to pick out. Rupert, on the other hand, offers an 'acquisition' account only in the sense of after-birth brain development, which is a merely brute-causal (non-rational/inferential) process.

15 No claims are being made presently about the actual perceptual space of human beings. The example serves merely to illustrate the kind of process that allows for the acquisition of simple
Imagine further the agent is designed to learn from its experiences how best to carve up its perceptible environment for its own purposes. The agent randomly generates syntactic strings to be used as representations. It might help to think of the agent as having two mental ‘buckets’ of arbitrary strings of syntax; in one bucket there are strings to be used as names for objects and in the other bucket there are strings to be used as names for kinds and properties. Suppose there is also a syntactic marker that stays with the symbol so that the agent can distinguish the object names from the property names (e.g., object names are in lower-case and property names are in capitals).

This machinery allows the agent’s perceptual system to present to the agent objects as having properties in terms of positions of the objects on its perceptual space. This is a straightforward sense in which the agent is able to compose its shape and color symbols together to entertain thoughts, like, OBJECT b HAS PROPERTIES C1 AND S1. Let each of the lower-case letters in Figure 4.2 be a randomly selected symbol from the ‘name bucket’ and assigned to objects the agent has perceived.

Now, imagine that the agent is able to detect patterns in its property space, as is done by commonplace clustering algorithms used in the Machine Learning branch of Artificial Intelligence (Mitchell, 1997). Clustering algorithms find sets of similar objects by measuring distances between them in a feature space.

Up to this point, our imagined agent is a crude version of what Concept Empiricists and Concept Nativists have always imagined human beings to be like. Both views would agree that such clusters lead to the acquisition of new concepts, and they would agree that the concepts that result are concepts.
Composite. They would disagree, however, about where the *lexical* concepts appear on this picture. Concept Empiricists would claim that most lexical concepts are such Composite cluster representations, whereas Concept Nativists would claim that most lexical concepts are themselves among the Primitive symbols that make up the agent’s innate perceptual space.

Of course, the Baptism of Mental Terms departs from both of these views. According to the BMT, lexical concepts, at least the ones for natural kinds, are **Acquired** via the detection of such clusters of representations already in possession, but are **Simple** in structure rather than being composed by the representations forming the cluster. On the BMT, the agent interprets clusters of objects in its perceptual space as an indication that the objects share a property that *explains* the observed clustering. The agent then forms a mental description to pick out that property. Once the (purported) property is picked out by a mental description, the agent then assigns a new **Simple** mental term, from its ‘bucket’ of property names, to the property there and then picked out. This way, a new simple name comes to represent that newly discovered property, the property of being an *apple*, as the case might be. This baptism is what initially determines the meaning of the concept, and the information making up the description is then stored as contingent information about things that fall under the concept. Figure 4.3 shows property names in capital letters that are associated with each cluster.

![Diagram showing names given to properties](image)

*Figure 4.3. Names are given to the underlying properties*

Once the clusters are detected and names are given to the properties responsible for them, the agent can represent the objects used to discover the property as having the property. For example, the agent can entertain the representations ‘g is an M’ and ‘c is an N’.

This machinery also allows new objects to be recognized as having the newly discovered property. If a new object is perceived as falling near enough to the objects that are taken to be Ms, and far enough from the other objects (where enough is decided, however arbitrarily, by the clustering algorithm), then the new object will be judged by the agent to be an M.

The new property representations are **Acquired** in that the acquisition process is a rational/inferential process. These new property representations are also **Simple** in representational structure even though they are **Acquired**. That is, the property name is neither syntactically nor semantically identical with the description. Notice first that ‘M’ is different in syntax from ‘C1 AND S1’. It is a
simple name that does not have representations as parts. Notice that ‘M’ is being used to name a newly detected property, the underlying property that is the inferred explanation of that observed pattern in the world. M was not acquired by the agent’s mere composition of its representations. It was the world that connected C1 and S1 in the agent’s experiences, giving the agent reason to form a new concept for the property that is responsible for the similarity in experience. Finally, this model allows the agent to think about blue apples and yellow apples, by simply composing M with other representations. This is because the meaning of the newly acquired concept is the property that is contingently picked out by the mental description, instead of being the description itself.

So the proposal is as follows. An agent assigns an arbitrary mental symbol to a property, when presented with a cluster of objects that have this property. The mental symbol refers directly to the thereby introduced property. In the vernacular of the philosophy of language, the agent ‘baptizes’ the property with the use of the mental symbol. We now turn to the thorny issue of the nature of such mental baptisms.

In the case of natural language, philosophers have appealed to an analogous process for the introduction of natural kind terms. For example, Hilary Putnam, when discussing how one can learn the meaning of the linguistic term ‘water’, seems to suggest a mental process very much like the Baptism of Mental Terms via reference fixing via similarities in a perceived sample.

Suppose I point to a glass of water and say, ‘this liquid is called water’... My ‘ostensive definition’ of water has the following empirical presupposition: that the body of liquid I am pointing to bears a certain sameness relation (say, \( x \) is the same liquid as \( y \), or \( x \) is the same \( L \) as \( y \)) to most of the stuff I and other speakers in my linguistic community have on other occasions called ‘water’ [pg. 225 in Mind, Language and Reality].

Putnam’s ‘ostensive definition’ here seems to require the same kind of mental work as is being proposed in the BMT. In picking out the kind, water, there is a presupposition that there is exactly one underlying kind that is common to a set of similar-looking samples.\(^{16}\)

Similarly, Scott Soames (2001, pp. 266-7), when extending Kripke’s account for the acquisition of linguistic natural kind terms writes:

According to [Kripke’s] account, the predicate is first associated by speakers with a kind –either ostensively or via a description. In the ostensive case speakers directly associate the predicate with a certain sample of individuals, which they presume to be instances of a single natural kind of a given type (e.g., a single substance or a

\(^{16}\) Notice that Putnam is careful to include ‘liquid’ in the presupposition, to ensure that the term ‘water’ gets hooked onto the property of being water rather than any of the other properties in the sample.
single species). In the Composite case, speakers employ a description that picks out a unique kind, often by appeal to contingent properties of the kind, or its instances.

Soames draws a somewhat artificial distinction between what he calls ‘reference-fixing by ostension’ and ‘reference-fixing by description’. I take both processes to be species of Composite reference-fixing, albeit ones that differ in that ostensive reference-fixing involves descriptions that contain perceptual demonstrative reference to some sample of individuals. Soames’ description of natural-kind term introduction by ostensive reference-fixing is analogous to the process of mental baptism of properties of the sort I have been discussing. A successful account of the nature of such a process involves provision of suitable reference-fixing descriptions.

4.2. Refining the Reference-Fixing Description
Let us call a description ‘The F proper if and only if the following two conditions hold:

(i) There is at least one F (existence)
(ii) There is at most one F (uniqueness)

A successful account of mental baptism will take the form of a proper description that will serve to fix the reference of the thereby introduced mental term, when perceptually presented with a sample of objects, clustered in the innate perceptual space. I now consider a series of candidate descriptions for the acquisition of APPLE.

Suppose that the sample of objects in a given cluster more or less shares the properties, \( p_1, \ldots, p_n \). The most obvious description the agent could use to pick out the reference is the following:

(Description 1) ‘The property of having the conjunction of properties \((p_1 \& p_2 \& \ldots \& p_n)\)’

According to the advocate of BMT who proposes Description 1 as the reference-fixing description, an agent who introduces the mental symbol APPLE when presented with a sample of apples, exploits a reference-fixing description such as (for example) ‘the property of being red and round’.

Description 1 is not, however, a plausible reference-fixing description. The property of being an apple is not even close to being extensionally equivalent to the property of being red and round. When we think about apples, we are thinking about that property; we are not thinking about red and round things.

A better approach is to say that when confronted with a set of sample objects, \( o_1, \ldots, o_n \), that are similar in appearance, there is a unique property that they all share. This motivates the following reference-fixing description, which is clearly ostensive in Soames’ sense:
Notice that Description 2 requires that there is a unique property that is shared by the objects in the cluster. The idea is that when presented with some apples, the agent infers that there is a unique property that the apples have in common, namely that they are apples. However, Description 2 will not work to pick out the right property. There may well be a unique property that the objects in the sample have in common (e.g. ‘being in front of me and being an apple and being red and being round and having slight indentations either on the top or the bottom, etc’). But this property is not extensionally equivalent to the property of being an apple. Presumably, few (if any) other apples have this conjunctive property.

The advocate of Description 2 could respond by restricting her attention to properties that are non ad-hoc in various respects, say simple non-conjunctive properties. On this view of properties, being red and being round are properties, but being the conjunctive property of being in front of me and having some indentations on the top is not a property. So understood, however, Description 2 fails even more obviously. There are too many properties the objects in the cluster have in common besides their being apples. They are all red, and all round, and since they are apples, they are also all sweet, and ripe, and edible. There is no unique simple property that is shared by all of the objects.

Perhaps a way to eliminate these extraneous common properties that the reference-fixing description be restricted to pick out only natural-kind properties. The properties of being red, round, sweet, and so on, are all simple properties of the objects in the sample, but, unlike the property of being an apple, they are not natural-kind properties. This leads us to Description 3 as a way of uniquely picking out appleiness.

(Description 3): ‘The natural-kind property that is shared by objects (o₁, &…& oₙ)’

At first, Description 3 seems to be a good candidate reference-fixing description to be used in the BMT. Two worries arise however. First, the insertion of ‘natural-kind’ in the description seems to suggest that the agent has to have the concept NATURAL KIND in order to acquire new concepts in this way. This means that the agent must either have NATURAL KIND innately or acquire it in some other way. For now, let us assume that this concept is indeed innate.

A more urgent worry is that Description 3 still fails to satisfy the uniqueness condition, in spite of eliminating simple properties that are not natural Kind properties. If in fact the members of the sample all have the property of being apples, they also have other natural-kind properties, like being fruit, and being organic. The linguistic version of this worry is well-known in the philosophy of language as the ‘Qua-problem’. Michael Devitt and Kim Sterelny discuss this problem in their (1999) textbook Language and Reality, when a natural-kind term is being grounded, or assigned a meaning.
It seems that the grounder must, in effect and at some level, “think of” the sample as a member of a natural kind, and intend to apply the term to the sample as such a member... The term is applied to the sample not only *qua* member of a natural kind but also *qua* member of a particular natural kind. Any sample of a natural kind is likely to be a member of many natural kinds; for example, the sample is not only echidna, but also monotreme, a mammal, a vertebrate, and so on [2nd ed, Pg. 91].

What is needed is a way to isolate the property of being an apple from all of the other natural-kind properties that are shared by the objects in the sample. For the Qua-problem, I propose that the agent must explicitly invoke the notion of explanation in baptizing the mental term. As was suggested above in the sketch of the model, the detected clustering should be an *indication to the agent that there is a property that is shared by the objects in the sample and is responsible for the similarity in appearance*. Of course, as with the notion of natural kind, including the notion of explanation in the description seems to suggest that the agent needs to have the concept EXPLANATION in order to acquire concepts through this process. For now, assume the agent has such a concept, so we can see the work that it can do. Indeed, let us try a description that uses the notion of explanation without the notion of natural kind, to see if it can do the jobs of both.

(Description 4): ‘the property that objects \(o_1, \& \ldots \& o_n\) have that explains their similarity in properties \(p_1, \& \ldots \& p_n\)’

This fourth description eliminates many of the non-intended properties. Although the set of objects in the sample of apples would have many other natural-kind properties, it is presumably only their property of being apples that *explains* the observed clustering over redness and roundness. The property of being a fruit would be explained by a larger cluster, one that includes blueberries, bananas, and peaches. Likewise, the property of being organic would be explained by an even larger cluster, one that includes fruit as well as other plant and also animal matter. See image 5.4 for a 3-dimensional space with six clusters within two clusters of clusters (i.e., super-clusters).

![Figure 4.4., super-clusters.](image)

Description 4 works quite well, but it doesn’t eliminate all of the unintended natural-kind properties. For, it only eliminates the super-ordinate natural kinds. Suppose that all of the apples in the sample were in fact of a particular variety of apple, like being a McIntosh. The property being a McIntosh apple is a natural-kind property, so that is the property that would explain the clustering of the sample of objects in the region of the property space they occupy. A set of observed apples that included other kinds of apples, like Granny Smiths, which appear greenish in color, would presumably form a super-cluster, over these two apple varieties.

To deal with this McIntosh apple problem, I submit that it is an instance of a more general problem that occurs frequently during human concept acquisition. I call it the Limited-Sample problem, and it is a problem that faces the cognitive agent, not the concept theorist. If a child is exposed only to a limited sample of apples, all of which are McIntosh apples, and their being McIntosh apples is what explains the similarity that leads the child to coin a mental term, the mental term thereby coined means McIntosh apple. Perhaps later, if the child encounters other varieties of apple and clusters them along with the original set, her mental term will come to mean apple. Alternatively, if the child sees a new set of apples, all of which are Granny Smiths, the child may form two separate clusters. If a third dimension of features, say a taste dimension, is added, resulting in a super-cluster of McIntosh and Granny Smith apples, the child may then pick out the property of being an apple via Description 4.

A more immediate problem for Description 4 is that without the notion of natural-kind explicitly involved in the description, it will not serve to pick out the property of being an apple. The property of having similar surfaces, having apple skins, is what explains the similarity in appearance among the reddish and roundish objects in the observed sample. The notion of a natural kind will have to be used in the reference-fixing description along with the notion of explanation. This brings us to Description 5.

(Description 5): ‘The natural-kind property that the objects \((o_1, \&\& o_n)\) have that explains their similarity in properties \((p_1, \&\& p_n)\)’

Description 5 does all of the needed work. The property of being an apple is the unique natural kind that all of the objects in the sample have that explains the clustering around being red and being round. If the agent can formulate such a mental description, she can thereby come to think about the property of being an apple. From this, the agent is a mere baptismal step from having a concept of apple that is both Simple in representational structure and Acquired from other representations.

Towards one final refinement, consider the following scenario. Suppose that a set of objects appears similar to an agent, under certain environmental conditions. If the agent takes the conditions to be

---

18 This observation is due to Adam Sennet.
19 This consideration was brought to my attention by Jeff King.
stable and notices a clustering over features, this is enough of an indication that the objects share a natural-kind property. It is enough of an indication, even if the objects don't in fact have the properties they appear, under those conditions, to have. This consideration generalizes to include the cases in which samples appear similar in normal conditions. The result is Description 6.

(Description 6): ‘The underlying natural kind property that the objects \((o_1, \& \ldots \& o_n)\) have that explains their similarity in appearances \((a_1, \& \ldots \& a_n)\) under environmental conditions \((c_1, \& \ldots \& c_n)\)’

I propose for now that we adopt Description 6 as the form for the reference-fixing description to be used in the BMT process for concept acquisition.

5. Further Defense of the BMT Model
What has been shown so far is that many of the historical and contemporary tensions between Concept Empiricism and Concept Nativism can be resolved by a model on which lexical concepts are at once Acquired and Simple, and that the Baptism of Mental Terms is one such model. This section summarizes the observations that have been made so far in support of the BMT model over the two major alternatives, Concept Empiricism and Concept Nativism, and then further observations are made in support of the view. In this section we also consider some anticipated objections and defend the view against them.

5.1 Responses to Arguments for Innate and for Complex.
First consider how proponents of the BMT can defend Acquired from arguments in favor of Innate. Recall the observation that humans seem to lock onto the right natural-kind concepts with very little experience, too little experience for learning to occur. This observation could be taken to suggest that many of our natural-kind concepts must be innate. Certainly, an acquisition process that involves the formation and testing of hypotheses would take quite a lot of experience, perhaps along with some kind of reinforcement feedback from members of the community who already have the concept.

In response, notice that the BMT acquisition process does not involve the formation and testing of hypotheses, but it is nonetheless a rational/inferential process, not a brute-causal one. With the BMT, the agent needs very little exposure to instances in order to acquire the concept. A single perceptual exposure is enough for the agent to infer that there is an underlying property responsible for the appearance of the object. That is, as long as the instance occupies a point on the agent’s perceptual space that is far enough from other clusters of points (far enough, as determined by its clustering algorithm). I suspect that the arguments from hypothesis formation and testing are the result of tangling together issues about word learning and concept acquisition. Word learning might well require the possession of a concept for the meaning being learned. The hypothesis might be, for
example, 'The word “apple” means APPLE'. Notice, however, that in the case of concept acquisition, all that is needed is some way of picking out the property, which can involve a description in terms of perceptual demonstratives.

Next, consider how the BMT can defend Simple from arguments that seem to support Complex. First, consider the argument from apparent analyticities. Recall that the tenet Complex seems to explain some intuitions of analyticities, like 'All cats are animals', but not others, like 'All cats are mammals'. The clusters of clusters, from Figure 4.4 show how the BMT deals with the analyticity intuitions without Complex. A cluster for CAT could be formed before a cluster of clusters for MAMMAL. Since all of the objects from the CAT cluster would be in the MAMMAL cluster, we get the apparent necessity of 'All cats are mammals'.

The other major argument for Complex that we must address is the one from Frege-like cases of the informativeness of identity statements involving co-referring terms. Empiricists explain the informativeness by claiming that the two mental terms have different meanings, suggesting that they must be different Complex representations.

With the BMT as a clear alternative, we can see how to deal with the informativeness without having to say that the terms inherit their meanings from their Complex parts. Instead, while under the RTM, we can turn to the syntactic difference in the simple mental terms – one term is F and one is G. There were two different descriptions used in baptizing the terms, and the agent had no reason initially to think that the descriptions picked out the same property. Notice that the BMT’s syntactic explanation for the informativeness is superior the one that the Nativist might give. A difference in syntax would explain the intuition that identities involving co-referential concepts are informative, but the Nativist doesn’t have anything to explain how or why an agent would have two Innate mental representations with the same referent. There doesn’t seem to be much room, on the Nativist's account, for an agent to come to have two different mental names for the same property. The BMT model, however, does make sense of there being two mental names. Even though the meanings of both names are the same external natural kind, the mental symbols come to have that meaning via very different reference-fixing descriptions. The two pathways of acquisition lead to the baptism of two different mental symbols.

5.2 Responses to Anticipated Objections
Let us now turn to some objections that might be raised to the BMT. One objection against Simple as the meaning-determining factor for concepts might be that the meanings of concepts seem to ‘drift’ over time. Consider the following scenario as an illustration. Young Kimberly sees some small black animals playing in her backyard. In fact, these animals are skunks. Kim uses her mental symbol J to think about the natural-kind property that these animals have that explains their similarity in appearance. The next day, Kim sees the group of skunks again, but this time, there are one or two

---

20 This case is due to John Pollock.
cats have replaced a couple of the skunks. Not noticing the differences, Kim continues to J to think about those animals. As days go by, the skunks are gradually replaced by cats, and Kim continues to recognize them under her concept J. As she grows up, Kim learns the word ‘cat’ and stores it as the English name for the property picked out by J. In this case, it seems, intuitively, that although Kim’s concept J picked out the property of being a skunk when first baptized, it has come to pick out the property of being a cat.\footnote{This sort of objection is familiar in the philosophy of language for shifts in the meanings of singular terms. Most famously, Gareth Evans discusses the case of ‘Madagascar’ in his (1973) classic “The Causal Theory of Names”. It is fair to say that it is still an unresolved problem for causal accounts of reference-fixing.}

This kind of drift in meaning is puzzling, and needs to be addressed in any full account of concepts. It is a problem for every theory. The BMT as stated only addresses the initial acquisition of lexical concepts. Perhaps an avenue for dealing with drift on the BMT is as follows. A mental description, of the form of Description 6, fixes the meaning initially. As the cluster shifts, the description that informs meaning-determination shifts as well. In effect, perhaps we can regard the symbol as being continuously re-baptized as the agent gathers data about its world.

Next, consider an objection to the strange result of BMT that agents don’t know what they are thinking about. When mentally baptism a term, the agent picks out some property via a mental description, but the agent doesn’t know which property is thereby picked out. How does the agent know, for example, that it is the property of being an \textit{apple}, rather than, say, that of being a \textit{pear}, that satisfies the reference-fixing description? This might be a reason to think that humans have to be innately connected to the properties that their concepts represent, for that is the only way to make sense of the intuitive access we have to the meanings of our representations.

This objection seems powerful at first, but on closer inspection we see that the observation is irrelevant. There is a phenomenology associated with many of our representations, what it’s like, for example, to be thinking about apple-ness rather that pear-ness. But this phenomenological difference should not be taken into considerations about the contents of our representations. Indeed, if we assume the Representational Theory of Mind, the question is almost impossible to frame. Notice that the only way for an agent to represent any property, according to the RTM, is to entertain its symbol for that property. Within this framework, it is not clear what it is for an agent to know which properties its own mental symbols represent, other than in terms of other symbols. The best an agent can do, ‘from the inside’, is use the symbols it has to think about the meanings of its symbols. There is no representation-independent access to the meanings of our concepts under the Representational Theory of Mind.

Let us consider now a set of objections that have to do with the terms ‘natural-kind property’ and ‘explanation’ being in the reference-fixing description. The way the account was stated, it looks to be committed to the agent explicitly formulating such a mental description during the concept-
acquisition process. If this is the case, then the agent must have the concepts NATURAL-KIND PROPERTY and EXPLANATION in order to acquire concepts in the way suggested by the BMT model. This means either that these concepts have to be possessed innately or that they have to be acquired via some other process.

For the issue of the reference-fixing description being *explicitly* represented by the agent during the concept-acquisition process, a couple of issues arise. First, it is not clear what it is for something to be explicitly represented. It seems clear enough that the description is not consciously represented in a way that is introspectible, since we don’t notice ourselves formulating such descriptions. But conscious representation is taken to be different from explicit representation. Perhaps all the work that needs to be done by the description can be done the mechanics of the agent. However, it is not clear to me what the difference is, and when considering how to implement such an agent the distinction only becomes more confusing. I will put this issue aside for now, and assume that in fact the description is explicitly represented.^[22]

This means that if Description 6 is the right description, and it involves the terms ‘natural kind’ and ‘explanation’, we have to consider the plausibility of the agent having these corresponding concepts either innately or by some other acquisition process that occurs before the BMT process is carried out.

For the current proposal, I shall take the concepts NATURAL-KIND PROPERTY and EXPLANATION to be innate. This postulation should not be uncomfortable for either the Empiricist or the Nativist. The Nativist is already comfortable taking the majority of our lexical concepts to be innate. For example, according to the Nativist, concepts such as CARBURETOR and APPLE cannot be acquired, and are therefore innate. It is difficult to see why these sorts of considerations do not also extend to the concepts NATURAL-KIND PROPERTY and EXPLANATION. If CARBURETOR cannot be acquired, then surely NATURAL-KIND PROPERTY also cannot be acquired. The Nativist is in no position to criticize alternative views for postulating innate possession of concepts.

The Concept Empiricist also should not balk at the innateness of NATURAL-KIND PROPERTY and EXPLANATION should likewise be comfortable for the Concept Empiricist. Classical Empiricists must postulate a great deal of innate structure. For example, Empiricists already allow for the innate representations that make up the perceptual space, as well as an innate mechanism for doing statistics over the perceptual experiences in order to acquire new concepts. In short, Concept Empiricists attribute to agents an innate mechanism for constructing theories about the world, given perceptual stimuli. It is not difficult to see how an innate mechanism for forming hypotheses would need to employ concepts of NATURAL-KIND PROPERTY and EXPLANATION.

Another objection worth considering is that the BMT account of concepts is incomplete, because it relies on all of these innate concepts, without giving an account of how these concepts come to mean what they do.

This objection is irrelevant to the present proposal. The BMT is only an account of the acquisition of some Simple representations from representations already in possession. Also, as I have just pointed out, even Empiricist accounts are incomplete in this way, as they assume that there are some innate representations without offering a meaning-determination account for the innate representations.

5.3 Positive Defense
As a final defense of the BMT, consider the following two positive results. First, consider a problem from the field of Artificial Intelligence (AI). A common objection to AI is that its systems’ symbols are at best meaningful to their human engineers and users, not to the systems themselves (e.g., Searle 1980). Even if a system were hooked up to the world in such a way that its purported concepts are causally related in the right way to what they represent, that would not be sufficient for the agent to be said to mean things by its symbols. With only a causal connection, the agent would be no more than a sophisticated thermostat or fuel gauge, as the AI critics say. What the BMT allows is for the agent to hook up its own symbols to the world as a rational reaction to patterns in it detectors. Concepts, on the BMT come to mean what they mean in much the same way as coins and words and most other representations do.

Second, notice how much power the view has when the simplified process articulated here generalizes in simple ways that make it psychologically plausible. Human beings presumably start with many more Innate representations than the ones for shape and color used in the example. Accordingly, human beings have a perceptual space with many more than two dimensions. Still, similarities between objects can be measured in the same way. The newly acquired concepts may in turn become further new dimensions in the similarity space, to be used in discovering further properties. The view may also extend to the acquisition of concepts through non-perceptual representations, such as through language. (To see how a computational dialog agent that uses the BMT can be designed to acquire concepts through language, see DeVault, Oved, and Stone, 2006).

6. Conclusion
In this paper I have shown that much of the debate between Concept Empiricism and Concept Nativism arises because of the Building Blocks assumption. In particular, I have shown that it arises from the assumption that if a representation is rationally/interentially Acquired from other representations, it must be Complex in representational structure, built up from those representations involved in its acquisition. When the Building Blocks assumption is abandoned, and a model is provided on which many concepts can be at once Acquired and Simple, we can see how the observations made by both Empiricists and Nativists can be accommodated in a single, unifying view.
In this paper I have limited the scope to the acquisition of natural-kind concepts, but I hope that my defense of the Baptism of Mental Terms shows that the view has sufficient promise to merit its application to other classes of concepts, such as concepts for artifact kinds and social kinds. The Baptism of Mental Terms should be, at least in hindsight, an obvious candidate for a theory of concepts.

References:


Locke, J. (1690) *An Essay Concerning Human Understanding*.


