Naturalization without Natural Kinds:

On the prospects of a unified theory of religion.

> From the propagation of religion, it is not hard to understand the causes of the resolution of the same into its first seeds or principles; which are only an opinion of a deity, and powers invisible and supernatural; that can never be so abolished out of human nature but that new religions may again be made to spring out of them by the culture of men as for such purpose are in reputation.
>
> -Thomas Hobbes, *Leviathan*, Part I, Ch xii, 23

Naturalization has dominated philosophy of mind for a generation. Philosophers have sought to explain perennial philosophic disputes such as the nature of consciousness, knowledge or intentionality in natural terms. Now that same movement has come to religious studies. Among others, Pascal Boyer, Harvey Whitehouse, Tom Lawson, and Bob McCauley seek to explain religion–or, more technically: religious concepts, rituals and behavior–in terms of the natural mechanisms of the human mind. The problem facing this movement is that there is general agreement that religion, unlike mental kinds naturalized by philosophers of mind, is not a natural kind. Thus, the nascent discipline of the cognitive science of religion appears to attempt the impossible: to naturalize an unnatural kind.

In this paper, I argue that this challenge is less impossible than it prima facie appears. Naturalization is the process of explanation through the specification of underlying mechanisms that realize the behaviors to be explained. Naturalists tend to think that naturalization begins by tallying all known instances of a phenomenon, and only then querying the underlying mechanism. This is not exactly right. Naturalization is a dynamic process, in which kinds that previously appeared to be natural are split, rejoined or discarded altogether.

The same is likely to be true of the study of religion. Religious beliefs and behaviors, which may never be explicable in terms of a single mechanism, can be explained by a disunified collection of natural and unnatural mechanisms, the exaggeration of each of which turns a nonreligious social or psychological behavior into a religious behavior. Unfortunately, these mechanisms also produce behaviors not deemed

religious, and hence cannot provide a basis for the unification of religion. This does not mean, however, that the beliefs and behaviors themselves are not natural kinds.

The argument proceeds in four sections. In the first, I relate three classic analyses of the concept of a 'natural kind' and argue that religion fails to satisfy any of the analyses. In the second, I consider the process of 'naturalization', in analogy to the history of cognitive science and the role of philosophy of mind in the formation of that field as well as the process of splitting and dissolving kinds. In the third, I review the theories of religion offered by Whitehouse, Boyer and Lawson and McCauley in light of the analysis of naturalization developed in section 2. In addition, I propose a new approach, based on religious attitudes, to understanding religion that complements these three. Finally, in the fourth section, I argue that it is possible to maintain the process of naturalization of religion, even without natural kinds, and draw out a few lessons for future progress in the cognitive science of religion.

First, a note on formatting. When referring to concepts, I will use all capitals, when referring to the word, I'll use single quotes, and to the objects themselves, I'll use the word. Thus, SQUIRREL designates the concept, 'squirrel' the word, and squirrel the particular squirrel about which we are speaking. By 'extension', we mean the set of objects covered by a particular concept. Thus, the extension of the concept SQUIRREL is the set of all squirrels. The extension of the concept THAT SQUIRREL is the particular, individual squirrel about which I am thinking

1. Natural Kinds

Tom Lawson, Bob McCauley and Pascal Boyer have all argued that religion is not a natural kind. Their arguments are well-known but, to my mind, unconvincing. The notion of 'natural kind' operative in these arguments is a rudimentary one. I argue that a more complex analysis of the concept yields the same result, but for different reasons. Before I can proceed with the argument, I must clarify some of the traditional philosophic analyses before applying them to the arguments put forward by Lawson, McCauley, and Boyer.

A natural kind, at its most basic, is a classification of events, entities, properties, relations, behaviors or states of affairs that group together naturally, rather than artificially. This preliminary notion raises the immediate problem of delineating between 'natural' and 'artificial'. Take, for example, a classic 'unnatural' kind: weeds. Suppose I am gardening one spring day, and I ask my wife "Is this a weed?" I am

not asking if this particular plant is unnatural–it obviously is a member of the set picked out by my concept 'natural'. Thus, I am not seeking an answer to a horizontal question, such as "Is this particular tree an elm?" I am seeking an answer to a vertical question, such as "is this plant a member of a species that is included under the super-concept WEED?" And the category 'weed' applies to species based not on properties intrinsic to them, but on these species' relations to their environment (i.e. 'invasive', 'alien') or their relation to humans (i.e. 'cultivated', 'despised'). Thus, a concept can be natural horizontally, but artificial vertically.

It may perhaps be better to distinguish between 'intrinsic kinds,' those kinds that group together solely on the properties of the individual members considered alone, and 'extrinsic kinds,' those kinds that group together based on the relations each individual member has to something else. The set of all 18-22 year olds hangs together intrinsically, in virtue of the fact that they came into this world between 18 and 22 years ago. The set of all students in my Critical Thinking class hangs together extrinsically, in virtue of the fact each of them has a certain relationship with me. I will return to this kind of delineation in a moment, but let me first consider some classic analyses of natural kinds found in the philosophic literature.

1.1. Projectible Properties

First, there is a tradition in the philosophy of science, usually attributed to W.V.O. Quine (1999a), that holds that natural kinds support 'projectible' properties while non-natural kinds do not: if a property holds of one member of the natural kind set, it *probably* holds of other members of that set. Thus, if 'mammal' picks out a natural kind, it is likely that all members of the set mammal will have the properties of any given mammal. There are two challenges to this approach. First, the duckbill platypus has a number of properties that are not shared by any other mammal, yet 'mammal' picks out a natural kind. Second, the property of being despised by gardeners is projectible to the set of all weeds, which is not a natural kind.

We can solve the problem by pointing out that the latter projectible property is a 'just-so' property. The statement that weeds are despised by gardeners contains a trivial truth. The property of laying eggs, which we may wish to project from the case of the duckbilled platypus is not a 'just-so' property. And the statement that mammals lay eggs does not contain a trivial truth. In fact, it contains a falsehood. One cannot falsify 'weeds are plants despised by gardeners', for if one found a plant that was beloved by gardeners, it would not be classified as a 'weed'.

On this account, we can distinguish between natural and artificial kinds, not on the basis of their relationship to human-classification schema, but rather on the 'just-so' nature of the projectible property associated with that natural kind. Properties projected from natural kinds are *likely* to hold for all the members of that kind. That is, they might *not* hold. Properties projected from artificial kinds hold *necessarily* for all members of that kind.

Religion does not admit of projectible properties. First, the term 'religion' or 'religious' can be applied to a system of practices, particular rituals, individuals, or sets of beliefs. A property of a system of practices may be completely inapplicable to a particular ritual in that tradition. A person may be deemed to be 'religious' if he or she participates in religious rituals, but not necessarily. And a set of beliefs likely shares few properties with the practices that correspond to it.

Second, one cannot predict that all and only members of the class *religion* will include ritual acts of self-sacrifice, meditative states, beliefs in supernatural beings, or any of the rest. For any one of these properties, scholars of religious studies can point out a litany of counterexamples. But one must be careful with this dialectic. Pointing out that the duckbilled platypus lays eggs does not make us think that the class *mammal* is any less of a natural kind. A single counterexample may undermine a particular projectible property, but it does not undermine the claim that the class constitutes a natural kind. If, on the other hand, there are no projectible properties, then we have reason for thinking that there is no natural kind. It would be premature to argue, on the basis of this dialectic, that the concept RELIGION should be discarded. For example, we shouldn't discard the concept RELIGION just because there are belief systems covered by that term that do not contain representations of the infinite. To discard the concept on this basis would be to assume, without argument, both that concepts are defined in terms of the necessary and sufficient conditions for their application and that concepts must correspond to natural kinds. Neither of these is true. The concept HEAP is useful, even if we cannot specify any specific projectible property that all and only heaps have (i.e. one that distinguishes between heaps and mounds). And it is very unlikely that HEAP picks out a natural kind.

1.2. Counterfactual Reasoning

A more sophisticated version of the Quinian approach is offered by Jerry Fodor and Zenon Pylyshyn , in their paper "How Direct is Visual Perception: Some reflections on Gibson's 'Ecological

Approach'". Fodor and Pylyshyn claim that natural kinds hold in counterfactual (unobserved) cases where mere generalizations do not. Consider the claim that all mammals have hearts and all mammals are born before 2010. Both claims hold for all observed cases, but only the former holds for all counterfactual cases as well. The claim is not that natural kinds actually hold in counterfactual conditions (as that is impossible), but that they can hold–i.e. that natural kinds support counterfactual reasoning while unnatural kinds do not. As Fodor and Pylyshyn say, the former "provides reason for thinking that there could be no mammals without hearts, while the later provides no reason for thinking that there could not be mammals born after 1982." (2002: 174)

Fodor and Pylyshyn's modal approach to natural kinds allow us further to rule out the kind of trivially true claims such as 'all creatures born on the fourth of July 2007 were born on the fourth of July 2007'. Such cases are not just trivially true, they are necessarily true: one cannot imagine a possible world in which that statement was false. Consider again 'all mammals have hearts'. That statement might, in fact, be false. It is unlikely, but possible. The former is impossible. Thus, natural kinds support reasoning about probably true but possibly false projectible properties, while unnatural kinds yield only necessary truths or fail to support reasoning regarding probably true projectible properties.

Does calling some behavior 'religious' provide reason for thinking that that behavior will continue in counterfactual conditions? The statement "all religious people engage in meditative behavior" provides me no reason for thinking that people who cease to engage in meditative behavior have likewise ceased being religious. It may be the case that meditation occurs only during part of a religious adherent's life. Or we can imagine a possible religion that refrains from meditation. An analogous case can be made for culturally postulated superhuman agents. To claim that all religions contain representations of culturally postulated superhuman agents provides me no reason for thinking that it is impossible that a culture without such representations would, for that very reason, be areligious. For example, suppose a culture that engaged in ritual behavior without technical motivation, yet lacked representation of superhuman agents. If this qualifies as a religion, then according to Fodor and Pylyshyn's counterfactual approach, the mere possibility of such a religion entails that 'all religious people have representations of culturally postulated superhuman agents' is false. In his paper "After the Naming Revolution: Joachim Wach's Unfinished Program," Alles argues that the category 'religion' is inference-poor, meaning that there is little one can

infer from the claim that some thing is a religion. His claim is consistent with this rejection of natural kinds as counterfactual-supporting generalizations.

1.3. Unity of Underlying Mechanisms

Consider the statement 'all mammals have mammary glands.' Is that statement not trivially true? In a sense, it is. But when the initial category was delineated, having hair, being warm-blooded and a host of other projectible properties were putative candidates for the defining feature of mammalianness. The process of classification is a dynamic one. Categories are often put forward tentatively and subjected to future revision. The process of definition is on-going, not a static stipulative process as assumed by Penner and Yonan (1972) but a dynamic dialectic between our categorical schema and the empirical facts we seek to categorize.

Natural kind terms are continually revised in light of new observation, experiments and models. Phlogiston was once considered a natural kind, the substance that was released in combustion. It is no longer, having been split and replaced by 'Carbon dioxide' and 'nitrogen'. However, the kind once called 'dephlogisticated air' may be the same as what we call 'oxygen' today (see, e.g., Kuhn, 1962: 54-55). Francis Bacon did not begin his investigation of heat with the stipulative definition '"heat is an expansive motion which is checked and struggling through the particles" (Book II, Ch XX: 135). He began his investigation of heat by cataloging all the phenomena that fit our common-sense notion of heat, all those closely related phenomena that do not, all those exhibiting degrees of heat, all those that reject heat, etc. Only then, through the process of cataloging, did he arrive at the definition of 'heat'. If one or more of these categorizations changes in light of empirical evidence, we would expect our definition to change as well. Naturalizing a concept requires a dialectic between our classifications of the world and the empirical facts of the matter.[1]

Both the Quinian tradition and Fodor and Pylyshyn's extension ground the identification of natural kinds in terms of the 'reasonability' of holding projectible properties onto other members of the set in

[1] The disagreement between those who believe science begins with stipulated definitions and those who believe that science begins with investigation is literally as old as science itself. Hobbes distinguished himself from Bacon on precisely these terms, arguing that lack of formal definitions was the cause of error in scientific investigation. It is worth nothing, however, that not only did the Royal Society model itself on the teachings of the *New Organon*, but Hobbes was never admitted to membership.

question. But neither view explains why one might find a projectible property reasonable and another unreasonable. It is my contention that what we find 'reasonable' grounds the dialectic between our definitions and the empirical facts of the matter.

Surface-level correlation fails to distinguish between what is reasonable and what is not. Being liquid is highly correlated with being H_20, but being H_20does not provide reason for thinking something is liquid. The converse is true of mammals having hearts. Having a heart is highly correlated with being a mammal, and we find it highly improbable that something without a heart would be a mammal. But that is explanatorily flaccid. The reason we find such a claim highly improbable is because we know that a heart is an essential part of an underlying mechanism of mammalian viability. Being liquid is not an essential product of being H_2O, at least without constraining temperature and pressure.

Counterfactual reasoning cannot proceed on mere correlation. If it were to proceed on correlation, there would be no reason to consider one correlated property as providing reason over another. Thus, counterfactual reasoning must suppose something deeper. Counterfactual reasoning of this kind supposes, implicitly or explicitly, an underlying structure or mechanism that realizes the natural kind in question. A single unified structure or mechanism suggests a natural kind, different structures or mechanisms suggest different kinds. If we were to discover a mammal that lacks mammary glands, it would play less havoc with our natural kind classification schema then if we discovered a mammal with internal systems built of silicon chips, wires and hydraulics.

"All mammals have hearts" supports counterfactual reasoning precisely because having a heart is required for the function of the underlying mechanism of mammalian viability. "All mammals have appendices" does not because an appendix is not required for the function of the mechanism of mammalian viability. Just-so predicates, such as 'all mammals are called mammals by English-speaking humans', are ruled out as they are not properties of internal, underlying structures.

The question of whether 'religion' is the product of a single, unified mechanism will have to wait until I further clarify what I mean by 'mechanism', and how we have come to distinguish mechanisms–and correspondingly natural kinds–in the history of cognitive science.