Richard G. Condon Prize
What’s Not to Know? A Durkheimian Critique of Boyer’s Theory of Religion
Sarah Henning Davis

Abstract This article examines Pascal Boyer’s theory of religion, in which he argues that religious beliefs are an inessential by-product of theory of mind functioning. I suggest that Boyer is emblematic of a new wave of thinking about religion that oversimplifies religion and theory of mind processes and in so doing misses something profound about human cognition. Through a reexamination of Emile Durkheim’s Elementary Forms of Religious Life, one of the classic (and classically misunderstood) works on religion, the present work elucidates what the recent theories are missing. Specifically, where Boyer’s arguments suggest that “mentally building a ‘god’ requires first building a ‘thinking agent,’” Durkheim suggests that mentally building a “thinking agent” may first require building a “god.” [religion, theory of mind, counterintuitive beliefs, society, sacred, profane]

Religious practices and beliefs pose an evolutionary conundrum for social scientists because it is difficult to justify their ubiquity and persistence given a “lack of apparent utility and their expense in terms of time and energy” (Lawson 2005:57). The evolutionary puzzle of religion is to explain “the continuing success of groups with strange requirements and seemingly inefficient prohibitions” (Iannaccone 1992:271). In recent years, many cognitive scientists and evolutionary psychologists have rejected traditional functionalist accounts of religion in favor of a “cultural selectionist” viewpoint that suggests that religion sticks around not because of any direct evolutionary benefit provided to human communities but, rather, because it is a byproduct of human cognitive processes (Atran 2002; Boyer 2001; Dennett 2006; Tremlin 2005; Whitehouse 2004).

Human cognition evolved in a social environment. Cooperation was essential for survival, as was the ability to identify who was cooperating and who was not. Pascal Boyer is prominent among scholars who argue that a byproduct of social cognition, and in particular of a theory of mind to which humans are innately predisposed, is the belief in supernatural agents and, consequently, religion. Rather than an evolutionary enigma, Boyer asserts that religion is a “predictable side effect of the human cognitive engine’s performance, just as heat is a predictable side effect of the combustion engine’s performance” (2005:86). Boyer is thought to be “instrumental in delineating the cognitive foundations of god concepts, showing how they connect with intuitive knowledge bases and activate key features of the social mind” (Tremlin 2005:75).
In the present contribution, I examine some of the foundational principles of Boyer’s theory of religion, particularly: (1) his focus on the high transmission rate of counterfactual and counterintuitive beliefs (such as invisible or all-knowing beings) and (2) the stability and seriousness of a particular subset of counterintuitive beliefs that activate theory of mind processes and account for “religion” as we know it.

I see Boyer as emblematic of a new wave of thinking about religion that oversimplifies spiritual commitments and theory of mind processes and, in so doing, misses something profound about human cognition. Much of what is missing from the new picture can be found in a reexamination of Emile Durkheim’s *Elementary Forms of Religious Life*, one of the classic (and, I believe, classically misunderstood) works on religion. Durkheim considers what he calls religion among humans to be a fundamental component of theory of mind rather than a byproduct of it. Where Boyer’s arguments suggest that “mentally building a “god” requires first building a “thinking agent” “(Tremlin 2005:75), Durkheim suggests that mentally building a “thinking agent” may first require building a “god.”

Most cognitive scientists read Durkheim’s *Elementary Forms of Religious Life* as functionalist theory in which religion is a force of social cohesion, important for group unity, in which what is being worshipped is society itself (see Wilson 2002). Durkheim’s arguments are rarely considered from a cognitive point of view. This is because the complexity of what he means by “society” is underappreciated. In the *Elementary Forms*, Durkheim weaves an argument in which he reveals the indispensable connection among complex cognition (representational through and in language), high-level sociality (understanding others as intentional agents), and what he describes as the religious tendency among human beings (giving objective, formal properties to aspects of human experience that are by definition unknown). Durkheim’s arguments may not solve the evolutionary conundrum, but they radically shift the question by making it impossible to take for granted what religion is.

**Boyer: Counterintuitive Beliefs**

To explain the universality of religion, evolutionary theorists have asked why religious practices and beliefs were part of the selection process. Typical answers, such as to answer metaphysical or existential questions, to establish moral order and social cohesion, or to allay anxiety all fall short at the level of evolutionary explanation.¹

Boyer, who does not hold religion to be an adaptation in its own right, argues from a selectionist viewpoint for the significance of cultural transmission suggesting that given the general properties of human minds, “certain types of representations are more likely than others to be acquired and transmitted, thereby constituting those stable sets of representations that anthropologists call “cultures”“ (Boyer 1994:391). Selective models “generally focus on the transmission processes as the main cause of recurrence” (Boyer 1992:32).
To explain the recurrence of religious ideas through this selectionist viewpoint, Boyer elaborates on what it is about religious concepts that makes them likely to be acquired and transmitted. To understand the particular “stickiness” of supernatural ideas, Boyer lays out some of the mechanisms of ordinary human cognition. He begins by explaining “domain concepts” or “ontological concepts,” including such categories as person, artifact, animal, inanimate natural object, plant (Boyer 2000:196). Ontological concepts are like a “conceptual skeleton” of human cognition. When a new object is categorized under a domain concept, the result is inevitably that one has intuitions or expectations about this new object. If, for example, an object is identified as belonging to the concept domain “PLANT,” one will expect that the identified object will grow, need nutrients, and be inanimate as opposed to the animacy of animal life. Boyer suggests that an ontological concept is like having a “mini-theory” (although not at a conscious level) of certain kinds of things in the world (2001:74).

All religious representations, according to Boyer, satisfy two conditions: “1) They violate certain expectations from ontological categories. 2) They preserve other expectations” (2001:71). He describes the violation as a “tag” added on to the expectations of the ontological category marking a violation but without undermining the category all together. For example, a tree that understands conversations would fall into the ontological category of PLANT. “Understanding conversation” is a tag that violates that category, but other inferences about the plant—that it grows from the ground and needs nourishment—would be retained.

Religious concepts likewise go against “normal” inferences: They describe, for example, “PERSONS (therefore with a body) without a body, NATURAL OBJECTS (therefore without psychology) with a psychology, PLANTS (therefore inanimate) with animacy” (Boyer 2001:74). It is their strangeness or oddity that makes these ideas memorable: “Their attention-demanding quality depends on the counterintuitive claims” (Boyer 1994:94). To balance the “attention-demanding” quality of the supernatural beliefs with the potential “learnability” of these beliefs, it is necessary to “strike a balance . . . take all intuitive ontologies as confirmed, except a few assumptions that are then explicitly described as violated in the case of the religious entity” (1994:407). Thus, it is not simply the violation that causes good recall but, instead, the “violation against a background of expectation” (Boyer 2003:127). The counterintuitive nature of religious beliefs makes them, according to Boyer, a memory optimum and accounts for their high transmission rate in human cultures.

**Stimulating Inference Systems**

What is important to human beings, because of their evolutionary history, are the conditions of social interaction: who knows what . . . who did what with whom. . . . Imagining agents with that information is an illustration of mental processes driven by relevance. Such agents are not really necessary to explain anything, but they are much
If supernatural beliefs take the forms they do in human populations because these forms are salient and particularly memorable, what suggests them to people in the first place? Boyer gives two types of explanation. One is that humans have these ideas because ordinary theory of mind functioning evolved to be extraordinarily sensitive and, thus, gets tripped into activation by objects that are not human. The second, and not necessarily mutually exclusive account, is that out of the host of supernatural beliefs that arise accidentally and have high transmission rates because of their optimal structure, there is a subset that suggests itself particularly powerfully because of the relevance to features of the social human mind.

Boyer argues that the human mind is a “bundle of inference systems, differently activated by different objects” (2001:116). These inference systems have evolved specifically to deal with particular “objects,” but once established they can be activated by a variety of objects. He compares this to the complex architecture of the auditory cortex that allows for the precise analysis of sound, necessary for human language. This vital evolutionary adaptation may have byproducts such as “the universal predisposition to remember and enjoy music” (Boyer 2001:150–152). In other words, Boyer suggests that prevalent cultural phenomena such as music (that do not suggest any adaptive benefit at face value) may exist because human cognition has evolved to make certain things salient, memorable, and even pleasurable. Music stimulates an inference system designed to be sensitive to sound, and, thus, it is likely to stick around.

If music is a byproduct of an evolved auditory cortex, of what is religion a byproduct? To understand inference systems as evolutionary adaptations, it is necessary to understand what kinds of needs humans had, as opposed to those of other animals, such that uniquely human inferencing systems developed.

Because humans evolved in a social context, survival entailed cooperation. Dependence on cooperation results in “commitment problems” because the expected benefit of cooperating depends on whether another will, in fact, cooperate. To cooperate, one must have some way of ensuring that what they produce for the benefit of others will be returned in kind, and not squandered. “It becomes crucial to be able to estimate one’s potential partner’s willingness to cooperate or defect” (Boyer 2001:7). Access to strategic information is vital for human survival, where strategic information is defined as “the subset of all the information currently available . . . that activates mental capacities that regulate cooperation” (Boyer 2000:204).

A host of features of human cognition are explained as solutions to the strategic problems that arise in cooperative living. As Scott Atran puts it, our brains are “wired to spot lurkers (and to seek protectors) almost anywhere” (2002:69). Perhaps the most important cognitive feature that evolved to “select mates, form friendships, and detect cheaters in social contracts” (Day 2005:91) is theory of mind. Theory of mind is the evolved human
capability to draw inferences about the mental states of others: their intentions, beliefs, and desires.

Boyer argues that those supernatural beliefs that are salient and serious enough to qualify as “religious” include “only a specific subset of the supernatural templates, namely those that activate our “theory of mind” expectations” (2000:6). Because our theory of mind expectations are of vital importance to our survival, the suggestion is that, just the way the auditory cortex may be activated by music, the theory of mind inference system may be easily activated by “objects” other than human agents. Thus, when cues that normally activate theory of mind expectations (like a human face) are encountered in nonhuman objects, that is, faces in the clouds (Guthrie 1993), we may attribute agency to these nonhuman phenomena.

The evolved drive to access other agents’ mental states to ensure cooperation makes humans predisposed to attribute agency. “God concepts are constantly underwritten by the intuitive ontology and inference connections that make agency detection and the theory of mind the powerful, efficient adaptations that they are” (Tremlin 2005:76).

A fundamental feature of theory of mind capabilities is the notion that “access to information is imperfect” (Boyer 2001:9). A precondition for attempting to regulate other’s access to information (including trying to trick them) is an understanding of the partiality of information. Many supernatural agents, unlike ordinary agents, are “full-access” agents who are imagined to have access to all strategic information (Boyer 2000, 2001). Boyer argues that this particular kind of supernatural agent is likely to be successfully transmitted, because the counterintuitive property that it contains (a PERSON that knows everything), is salient given the importance of strategic information in human groups. “Imagined religious agencies are in general credited with good access to information . . . what religious agents know is what matters to social interaction in a human group” (Boyer 2000:11). Thus, the domain concept violation that characterizes these supernatural agents creates an agent that individuals would be particularly interested in. “Although gods, ghosts, and spirits resemble any other run-of-the-mill intentional creature, what is noteworthy about these agents is that they routinely have full access to the strategic social information that we limited-access actors so desperately want” (Day 2005:93).

Boyer has argued that religious beliefs have a high transmission rate because of their counterintuitive structure, violation against a background of expectation, which is a memory optimum for the human mind. God concepts arise, according to Boyer, as spandrels because of glitches in the social mind.

**Rereading Durkheim**

Boyer and Durkheim are writing almost a century apart, and Boyer is not in direct dialogue with Durkheim; nonetheless, it is significant that they share an interest in theory of mind.
(although Durkheim would not have called it that) and the counterintuitive nature of religious beliefs. Contemporary evolutionary theorists of religion begin their study with the notion that they must explain ubiquitous, energy-consuming religious behavior in a way that shows either that it increases evolutionary fitness or that it is an offshoot of processes that are themselves evolutionarily beneficial. It behooves anthropologists to return to Durkheim because his analysis requires us to think carefully about what it is that we are calling religion. If, as Durkheim makes clear, certain necessary features of human cognition are unthinkable without something that looks like religion, the starting point and framework for analysis change. To reframe in this way, however, one has to resist assuming that it is known from the outset exactly what religion is: “If religious faith was not born to put man in harmony with the material world, the injuries which it has been able to do him in his struggle with the world do not touch it at its source, because it is fed from another” (Durkheim 1915:83).

Theory of Mind and the Totem

Durkheim agrees with Boyer that access to others’ intentional states is always partial. However, where Boyer uses the fact of partiality to argue that an all-knowing agent would have particular appeal to a human mind, Durkheim looks at what such partiality suggests about human cognition in general. Durkheim’s arguments force the question: How is understanding of theory of mind complicated if part and parcel of knowing another as an intentional agent is knowing that we can never know his or her intentions fully, beyond the shadow of a doubt?

For example, if I run into Joe and I want to know if he is a cheater, I have to consider what he is up to. Constitutive of that consideration is the notion that he is up to something that I cannot access in its immediacy (which is why I am so intent on uncovering it in the first place). How do I represent this “inner state,” which I would like to discover, to myself, when it is precisely that which I do not know? There is a “something” that I am attending to, but that thing is defined only by its not yet being known. This “thing” that Joe possesses, of which I must be aware if I am to make useful assumptions—inferences or enquire further about it, is unlike other “things” in my environment. It exists for me, not like an animal, plant, or inanimate object but as a question, a “what?” To think about the “what” of which his inner state consists, I must posit it as a state to be known. To do this, I must give, what is actually an unknown, a positive form. I need a placeholder for it.

Durkheim’s investigation of totemism in The Elementary Forms of Religious Life is a study of what I have called the “placeholder” that allows humans to represent a state of being that is defined by its opacity. Because “we are unable to consider an abstract entity, which we can represent only laboriously and confusedly . . . we can not explain it to ourselves except by connecting it to some concrete object of whose reality we are vividly aware” (Durkheim 1915:220). To cogitate about another’s inner state, man conceives of it “under the form of an
animal or vegetable . . . in a word, of a visible object. This is what the totem really consists in: it is only the material form under which the imagination represents this immaterial substance” (Durkheim 1915:189).

Built into totemic belief systems is the notion that Joe is both what we see immediately, a man with two arms and legs, and so forth, and that which we cannot see immediately—his inner state. His inner state is represented by his totem, as, for example, a fox an eagle, a turnip. “Man believes that while he is a man in the usual sense of the word, he is also an animal or plant of the totemic species . . . the name [the totemic name] . . . is a part of the being, and even something essential to it . . . each individual has a double nature: two beings coexist within him a man and an animal” (Durkheim 1915:134). It is experiencing “Joe” as “double”—both a physical being and something beyond that physical immediacy—that allows me to wonder what he is up to.

It is this process—of giving positive form to a negative concept—that Durkheim considers the religious tendency among humans. Why must it be “religious” and not just a cognitive process that evolved in order for human minds to function socially? For Durkheim, religion and this evolved cognitive process are one and the same. As soon as that placeholder emerged, regardless of whether it was a totemic fox, the word what?, or any other representation or symbol, the world was split in two: the immediate was distinguished from the beyond and we were able to wonder what was going on inside the heads of other people. “We have religion as soon as the sacred is distinguished from the profane” (Durkheim 1915:182).

Unlike Boyer, who would suggest that the notion of a totemic beet, for example, is counterfactual (how can a man be both a human and a beet?), Durkheim reveals that the counterfactual nature of this belief is in fact rooted in a necessary cognitive move that lies at the foundation of theory of mind, giving positive form to an unknown that is defined precisely by its having no identifiable positive form. Rather than a cognitive malfunction, Durkheim is arguing that something like totems are required for human cognitive functioning.

The “beyond” that the placeholder allows human minds to consider is not a supernatural beyond, rather it is the realm of human experience of things that are not immediately transparent, primarily others’ inner states. However, the fact that the beyond is not mystical or supernatural does not mean that it is straightforward. In fact, because another’s intentionality, as just discussed, is necessarily veiled to a certain extent, it is, by its very nature, enigmatic. It is never simply known; there is always a question mark that accompanies it. Consequently, there will always be a contradiction embedded in our representations of that “beyond,” for it is precisely its undefined character that requires that the human mind concretize it in the form of a placeholder, but by doing so its inherent and inevitable elusiveness is at risk of being lost in translation. The very cognitive process that makes it possible to consider what other people are up to also makes representations of their inner
states deeply problematic. According to Durkheim, it is out of this inevitable, problematic character of theory of mind processes that the mystical qualities of institutional religion arise.

Durkheim discusses the inevitable “secondary error” (1915:226) that occurs when inherently negative concepts are given positive form. The secondary error is the slippage that goes on when humans take the object that, as a symbol, is meant to be a placeholder for an unknown, and reify it. We treat the placeholder, for example a totemic fox, as though it is not just the “stand-in” for the unknown, but the unknown itself. A totemic emblem becomes understood not as the placeholder for one’s inner state but, rather, as the bearer of it. The emblem is “loved, feared, respected; it is to this that we are grateful; it is for this that we sacrifice ourselves . . . one loses sight of the fact that the [totemic emblem] is only a sign, and that it has no value in itself, but only brings to mind the reality that it represents; it is treated as if it were this reality itself” (Durkheim 1915:220).

If theory of mind processes require that a positive form be given to a phenomenon that has no positive form, revealing a counterfactual premise deeply embedded in our sociality and complex cognition, how does this contradiction surface in human societies? Durkheim suggests that institutional religion is one way. By worshiping the concrete form of the placeholder as the embodiment of all that is unknown, the counterfactual at the source of social cognition gives way to a belief in a concrete embodiment of all that is enigmatic. The immediate—beyond split, necessary to understand other’s intentions, becomes a split between the profane and the sacred.$^2$

While institutional religion, rites, and rituals may be a byproduct of the contradiction embedded in the social mind, the contradiction itself, which exists in all human societies, as it is at their foundation, is what Durkheim considers to be the core of the religious tendency in humans. This contradiction is not a byproduct of the social but the very source of its functioning. It is through this lens that

We are . . . able to understand how the fundamental truth contained in religion has been able to compensate for the secondary errors which it almost necessarily implies, and how believers have consequently been restrained from tearing themselves off from it, in spite of the misunderstandings which must result from these errors. It is undeniably true that the recipes that it [religious tendency] recommends that men use to act upon things [rites and rituals] are generally found to be ineffective. But these checks can have no profound influence, for they do not touch religion at its fundamentals.

[Durkheim 1915:226]
its very nature, contradictory. Humans posit a state that exists, which theoretically could be totally known. However, humans also know that each person is fundamentally blocked from having access to the completeness of that state; if this fundamental block were ever removed the enquiry would lose relevance. Humans are compelled to consider each other’s inner states because these states are both known and unknown. What one knows about others is that they possess an inner state, what one cannot be sure of is exactly what this state consists of in any given case. This creates a kind of vacuum of knowledge that draws individuals together. Durkheim’s argument is that society is bound together by the “unknown” aspect of humans rather than the known. Thus, rather than being bound by common, transparent humanness, common physical features, needs, and so forth, Durkheim argues that people are bound by that which they share that is not transparent. It is not knowing what one another are up to, but knowing that each person is up to something, that makes each individual relevant to each other in a powerful and social way. In a totemic belief system, if someone says he or she and fellow clan members are crows, “he [or she] does not exactly mean to say that they are crows in the vulgar empiric sense of the word, but that the same principle [their common opaque humanity] is found in all of them . . . which is thought of under the external form of the crow” (Durkheim 1915:189). Durkheim argues that “it is to this common principle that the cult is addressed . . . anonymous and impersonal . . . found in each of these beings but not to be confounded with any of them” (1915:188).

As I have suggested, Durkheim argues that this common, compelling unknown, requires a placeholder to exist cognitively for humans. Humanity’s unique cooperative and social unity is born with the placeholder. For Durkheim, the totem (the placeholder) and the clan (in his view the first form of human society), emerge simultaneously. They “mutually imply each other” (Durkheim 1915:167). Totemism, he suggests, is “inseparable from social organization on a clan basis. Not only is it impossible to define except in connection with the clan, but . . . the clan could not exist . . . without the totem” (1915:167). For Durkheim, human society, unique in its capacity for cooperative living and complex communication could not exist without the placeholder: “the primitive family organization can not be understood before the primitive religious beliefs are known; for the latter serve as the basis for the former” (1915:106). He argues that clan unity does not come from blood but “solely from their having the same name and the same emblem . . . in a word from their participating in the same totemic cult” (1915:167).

When Durkheim argues that “God and society are one” (1915:206), he is not, as is often suggested, making a simplistic functionalist argument in which humans get together in cooperative society and then create a god that stands for that society and worship it to ensure social cohesion (Wilson 2002). Rather, in a more careful argument he is suggesting, like contemporary cognitive scientists, that human society emerges when humans are able to conceive of each other as intentional agents. What often goes undertheorized in the current literature (Atran 2002; Boyer 2001; Dennett 2006; Tomasello 1999), however, is Durkheim’s notion that understanding others as intentional agents requires that we understand the partiality of human knowledge, which he reveals to mean that humans must posit a
totality that can never be accessed. This totality, precisely because it can never be accessed, is cognitively diffuse and difficult to manage. To consider it, humans must give it a positive, transparent form. This form, which veils the formless nature of the phenomenon that it is meant to represent, Durkheim suggests, is God. The contradiction embedded in the form is both the glue that pulls people together and also the source of religious mysticism.

The Counterintuitivity of Language

Boyer links god concepts to theory of mind processes because he wants to explain why humans would attribute intentional agency and exaggerated theory of mind capabilities to things that have none. He sees that the same process that humans use to understand each other is somehow being applied to nonhuman phenomenon and exaggerated in these cases. Boyer’s argument seems to be straightforward. Humans possess theory of mind and in the case of religious beliefs are somehow misapplying or extending this cognitive ability, thus his project is to explain the misapplication—extension.

Durkheim, however, does not assume that the human ability to understand others as intentional agents is straightforward. He picks apart what is involved in this cognitive process and shows that in order for humans to conceive of each other’s intentional states, they have to be able to give form to something they do not yet know, and will never know fully. Given the enigmatic character of humanity’s intentional state, Durkheim suggests that the placeholder that best captures it will be one that can indicate the simultaneously known and unknown character at the heart of the concept. What kind of placeholder can simultaneously have discoverable content and always point beyond that content, reminding us that the placeholder is not identical to the state it indicates? The placeholder must be finite and infinite, known and unknown, particular and unparticular.

Durkheim suggests that the placeholder that allows humans to put the cognitively complex notion of an “unknown” into consciousness is representation itself. Rather than choosing an object in the natural world, for example a pig, to stand for an individual’s inner state, in totemic belief systems it is rather the etching of a pig that acts as the placeholder. Durkheim shows that the representations of the totems are considered more sacred than the animals they represent: “[Totemic objects] are distinguished from profane things of the same sort by only one mark of particularity: this is the totemic mark . . . drawn or engraved upon them. So it is this mark alone which gives them their sacred character. . . . The representations of the totem are . . . more powerful than the totem itself” (Durkheim 1915:133).

Durkheim shows that built into the very structure of human representations is the dualism that that placeholder is born to indicate. Consider, for example, a drawing of a pig. As Boyer would agree, in so far as it is meant to represent “pigness,” the representation must adhere to the ontological expectations of a pig: it has a physiology, it is animate. However, the drawing...
of the pig also violates those expectations because the drawing is just lines, not in fact a real live pig. It does not have a physiology; it is not animate. This drawing, and all representations for that matter, both are and are not the things they represent. This duality is at the heart of language in which words, written or uttered, are and are not the things that they indicate.

Thus, according to Durkheim, the fundamental dualism embedded in our notions of others as intentional agents, emerges not only in the secondary errors of institutional religion and in the compelling force of society, but in the very structure of our cognition, through representation and language. Durkheim turns our lay understanding of language on its head and reveals that rather than being a straightforward system of reference for the known world, human representation and language arise so that human minds can conceive of the unknown. Subsequently, language is used to carve up and categorize empirical reality, but Durkheim's radical point is that the compulsion for humans toward language was not to refer to that which was immediately accessible, but to that which without reference was cognitively unmanageable. It is for this reason that Durkheim argues that the totem not only enables theory of mind and marks the birth of society but also is “the first form of . . . a written language” (1915:127).

From a Durkheimian point of view, although Boyer rightly points out the counterintuitive nature of religious beliefs (the violation of ontological categories against a background of expectation), he fails to understand the genesis of this counterintuitivity. Boyer fails to realize that his interest in concepts that go against “normal” inferences is a fundamental component of all representation. Because Boyer does not see the interconnectedness of religion and cognition, he analyzes the counterfactual nature of religious beliefs in a vacuum, rather than seeing it as an exaggerated (and, therefore, revealing) version of a tension that is at the heart of all language.

**Consequences**

Durkheim argues that if we understand the religious tendency among humans in the way he outlines, and the inevitable secondary errors that it provokes, “religion ceases to be an inexplicable hallucination and takes a foothold in reality” (1915:225). By looking carefully at what cognitive processes are necessarily implied by theory of mind, Durkheim’s understanding of religion makes a point that is radical for today’s social scientists: that “the believer is not deceived” (1915:225). “The images out of which it [religion] is made are not pure illusions . . . they correspond to something in reality” (1915:227). The totemic emblems refer to the common, enigmatic inner state of human beings. “It is true that he is wrong in thinking that it is the work of a power in the form of some animal or plant. But this error is merely in regard to the letter of the symbol by which this [state] is represented to the mind and the external appearance which the imagination has given it, and not in regard to the fact of its existence” (1915:225).
Boyer’s theory of the cultural transmission of religion attempts to explain what it is about counterfactual beliefs in the supernatural that makes them stick around despite the fact that they don’t correspond to anything in reality and are a drain of time and energy on the societies in which they exist without any explicit evolutionary benefit.

Durkheim’s work shows that these premises for a study of religion are short sighted. By pushing past our ordinary understandings of religion, Durkheim presents a probing look at the ways in which theory of mind, the clan, and totemism are mutually constituted. The *Elementary Forms of Religious Life* reveals over and over again that at the base of these defining features of human nature is the ability to consider seriously what something is before it is known entirely, which means that humans form some “idea” of that which they would like to know before they know what it corresponds to in reality. Humans are, in fact, forming an idea of something that cannot be found in empirical reality. It is not possible to point to an “unknown” in the same way that one can identify a tree or a pig.5

Rather than immediately assuming that beliefs in things that cannot be found in reality (supernatural beliefs) are cognitive malfunctions, Durkheim reveals that the human mind has to produce an idea of an “unknown” with formal properties to begin enquiring about that unknown. Thus ideas about things that cannot be found in reality are not aberrations, but rather the foundation for subsequent cognitive enquiry.

Because Boyer takes ordinary understandings of religion at face value, this powerful insight is lost. Understanding religion for many contemporary theorists, such as Boyer, Atran, and Dennet, has become a project of solving an evolutionary conundrum that exists only because religion and theory of mind themselves have been misunderstood. These misunderstandings not only diminish the analytic depth of our studies of religion and cognition, they also encourage a simplistic understanding of religious groups as “irrational” or as falling prey to cognitive malfunction. This is not only inflammatory at a sociopolitical level, but it is also inaccurate. Durkheim does not deny that all sorts of religious beliefs, rites, and rituals mistakenly attribute power to deities because the concreteness of the placeholder is overly emphasized. However, if the compulsion to concretize the unknown is understood to lie at the core of all of human cognition, religion begins to look less irrational and reveals itself to be a foundational, if volatile, aspect of humanity.

SARAH HENNING DAVIS is a Doctoral Candidate in Anthropology, Emory University.

Notes

1. See Atran 2002 for an evolutionary critique of functionalist explanations of religion.

2. Anthropologists could look at different religions and consider the ways in which they deal with the secondary error that emerges in the process of creating placeholders for the “unknown.” Max Weber’s study of Protestantism (1958) is in effect, just such a study. He suggests that in the move from Catholicism to Protestantism, God becomes more elusive,
and the emphasis on the unknowable aspect of the placeholder more prominent. Rather than considering God a manifest embodiment of the unknown, identifiable through icons and the priest’s word, Weber argues that the Protestant notion was that the unknown exists in each individual and how much God is with any individual can never be known for sure. In Protestantism, the placeholder for the unknown, rather than a fixed icon, becomes personal output, the concrete form that one’s energy takes in the world through work. Thus each individual’s drive for productive output, and the almost obsessive preoccupation with productivity, prosperity, and wealth was, Weber argued, not a form of greed, but a novel way in which the counterfactual religious tendency at the heart of social cognition was emerging in society.

3. This argument has much in common with Gregory Bateson’s discussion of play in *Steps to an Ecology of Mind* (1972) in which he argues that it is only in the human ability to express “not” that we are able to express indicatives (“is”). He suggests that before humans can express “not,” they cannot distinguish between seems–is. Durkheim argues, similarly, the use of placeholders (representations) among humans arises to stand for something that is “not” present, which then allows humans to use them to stand for something that is present.

4. Durkheim’s thought, in this regard, is clearly influenced by Jean-Jacques Rousseau’s understanding of the social contract (Rousseau 1978), in which Rousseau rejects a Hobbesian notion of society where humans enter society to escape the dangers of a state where each is driven by self-interest. Rousseau’s arguments are taken up by Talcott Parsons (1937) when he claims that a theory that posits human nature to be essentially self-interested cannot account for the key moment when an individual decides it is worth it to relinquish immediate passion (a piece of meat she or he is hungry for, a sexual partner, shelter) and enter into a social contract. To explain this moment, Parsons and Rousseau suggest that utilitarianism must posit (although utilitarianists never address it) that there is some kind of already existing relationship–contract in place, not based on rational self-interest, which allows this decision to be made. The fact that utilitarian theories, which may explain human behavior once society is formed, cannot account for the origins of the society in their own terms makes Parsons declare utilitarianism inherently unstable (1937:162). It is the same kind of instability that Durkheim is revealing that lies at the base of referential theories of language, which may make sense once language gets going, but on their own cannot account for the human capacity to understand that one thing can stand for another.

5. One might suggest that there are many things that humans encounter that they do not “know” fully. What does a peach taste like? How cold is the water? Such not knowing might cause a human (or other animals) to approach these experiences tentatively. The difference between these experiences and theory of mind is that unlike plunging into the water or biting the peach, humans *never* simply uncover the inner state of another. Each person is constantly in the process of assessing and reassessing. Knowing others is always as much about not knowing as it is about knowing. It is this “not knowing” (which results in the preoccupation with other human beings, namely sociality) of which humans must be constantly aware if they are to conceptualize other’s intentional states. It is the placeholder that serves to represent this “not knowing.”

References Cited

Atran, Scott

Bateson, Gregory

Boyer, Pascal


Day, Matt

Dennett, Daniel C.

Durkheim, Emile

Guthrie, Stewart

Iannaccone, Laurence R.

Lawson, E. Thomas

Parsons, Talcott

Rousseau, Jean-Jacques

Tomasello, Michael

Tremlin, Todd

Weber, Max

Whitehouse, Harvey

Wilson, David Sloan