Religion: Spandrel of the Brain?

The Evolutionary Psychology of Religion

Jorn Bruggeman
July – September 2003
Religion: Spandrel of the Brain?

The Evolutionary Psychology of Religion

Jorn Bruggeman
July – September 2003

Supervision: Cor Zonneveld

Cover: photograph of spandrels, taken from Carter (2002). Spandrels – the tapering triangular spaces formed by the intersection of two rounded arches at right angles – are necessary architectural by-products of mounting a dome on rounded arches. Yet, many spandrels have been so beautifully and elaborately painted upon that one is tempted to view them as the starting point in any analysis, as the cause in some sense of the surrounding architecture. But this would invert the proper path of analysis. Architectural constraints created spandrels; only subsequently could they serve as canvas (Gould and Lewontin 1979). Similarly, I argue in this study that religion does not exist to serve a specific cause: it is not an adaptation. Rather, it can be regarded a by-product of our adaptive cognitive architecture.
# Table of contents

Table of contents ....................................................................................................................... 2  
Abstract ..................................................................................................................................... 3  

## 1 Introduction .......................................................................................................................... 5  
  1.1 Religion in human evolutionary history ................................................................. 5  
  1.2 Religion as adaptation ............................................................................................ 6  
  1.3 Behavior as product of a modular brain ................................................................. 7  
  1.4 Religion as by-product ............................................................................................ 8  

## 2 On human nature .................................................................................................................. 9  
  2.1 Object detection ....................................................................................................... 10  
  2.2 Inference modules ................................................................................................... 13  
    2.2.1 Intuitive physics ............................................................................................... 13  
    2.2.2 Intuitive biology .............................................................................................. 14  
    2.2.3 Intuitive psychology ....................................................................................... 16  
  2.3 Social systems: kin, friends and communities ........................................................ 18  
    2.3.1 Parental care ..................................................................................................... 18  
    2.3.2 Kin selection ..................................................................................................... 20  
    2.3.3 Reciprocal altruism .......................................................................................... 23  
    2.3.4 Indirect reciprocal altruism: communities ......................................................... 25  

## 3 The naturalness of religion .................................................................................................. 29  
  3.1 Pathways to supernatural concepts ........................................................................... 29  
    3.1.1 Over-detection of agency ................................................................................. 29  
    3.1.2 Misfortune and social intelligence ................................................................... 30  
    3.1.3 Moral intuitions ............................................................................................... 31  
    3.1.4 Death ............................................................................................................... 32  
  3.2 On supernatural agents ............................................................................................. 33  
    3.2.1 What supernatural agents are like .................................................................... 33  
    3.2.2 Why supernatural concepts seem real .............................................................. 33  
  3.3 Evolution of religion ................................................................................................... 34  

## 4 Evaluation .......................................................................................................................... 35  
  4.1 Why religion is not (only) about .............................................................................. 35  
    4.1.1 ...existential questions ..................................................................................... 35  
    4.1.2 ...morality ......................................................................................................... 36  
    4.1.3 ...mystical experience ..................................................................................... 36  
    4.1.4 ...rituals ............................................................................................................ 37  
  4.2 So where does this leave us? ................................................................................... 39  
    4.2.1 To the critics ..................................................................................................... 39  
    4.2.2 Benign or evil? .................................................................................................. 39  
    4.2.3 Religion and science: distant cousins? ............................................................... 40  

## 5 References .......................................................................................................................... 42
Abstract

During the last decades, research into the biological origin of religion has gained considerable momentum. Religiousness has been shown to be heritable – though only to a moderate extent – which has prompted various scholars (e.g. D.S. Wilson, R.D. Alexander) to argue religion is an adaptation: a genetically inherited trait that increases fitness. However, as Stephen J Gould has forcefully argued, not all inheritable traits need be adaptations; in various ways traits can emerge and persist in evolution without increasing fitness per se. In this literature study, I explore the possibility that religion is not an adaptation, but exists for different reasons.

Non-adaptive traits can persist as by-products of adaptive traits. A number of scholars (P. Boyer, L. Kirkpatrick, S. Atran, S.J. Mithen, S. Guthrie) have argued religion qualifies as a by-product of cognitive hardware. In their view, religious thought and behavior is produced by various brain mechanisms that in themselves are adaptive. These were preserved in evolution because they solved particular (non-religious) problems faced by our ancestors. By-product theories of religion provide an interesting and plausible alternative to their adaptationist counterparts: they are capable of explaining the full spectrum of religious manifestations, without invoking any function for which they were selected.

Five million years ago, the evolutionary paths of man and chimpanzee separated: our last common ancestor eventually gave rise to a distinct human ancestor – *Australopithecus* – and an ‘ancestral’ chimpanzee. In the period to follow, the ape-like *Australopithecus* came to evolve into the modern *Homo sapiens sapiens*; selection pressures over the last five millions years gave rise to all traits commonly considered uniquely human (e.g. language, culture). What were these selection pressures like? Many have argued that our ancestor most prominently was under great pressure to organize: formation of ever-larger groups would have been indispensable for coordinated hunting- and gathering expeditions, and defense against predators (animals, but also other human groups).

To enable increases in group size, our brain has become strongly focused on other humans: it ascribes them beliefs, intentions and emotional states, feels their pain, shares their moral codes, and watches closely if others adhere to such codes. In fact, the human brain has become so strongly focused on humans that it quite easily perceives humans when there are none: any inexplicable phenomenon, any single or remarkable event is ascribed to human agents. In addition, the new cognitive hardware for active reasoning about other humans tends to over-activity: it fails to shutdown when observing familiar corpses (thus leaving the brain with the concept of a bodiless human agent), and prefers to represent the subconscious self as beliefs and intentions of others. As a result, the brain cannot help but observe an environment flooded with (signs of) powerful, invisible supernatural agents. Given the means through which it does so, it is not surprising that the brain holds quite elaborate assumptions on the nature of such agents: they are thought powerful, invisible or acting remotely, and involved in social exchange (and hence, to a certain extent manageable – one could achieve something by engaging in social interaction with the supernatural).

Summarizing, the brain inadvertently develops the notion of relevant supernatural agents: certainly an ideal starting point for religious thought and practices. Five million years
of specializing in social interaction rendered a well-adapted brain that fluidly – yet inadvertently – constructs its own religions.
1 Introduction

Religion certainly comes natural to humans: throughout the world, one cannot pinpoint any culture without religious beliefs or practices. Its manifestations vary greatly: from belief in dead ancestors controlling current events, collections of myriads of specialized, bickering deities up to monotheistic varieties with elaborate doctrines. Yet, all share particular characteristics (most typically, a belief in supernatural beings) that warrant the title ‘religion’. But what makes religion so natural that it is to be found in every human group?

1.1 Religion in human evolutionary history

To answer this question, one can begin with a focus on the origin of religion: was there a time in human history when human societies lacked all forms of religion? If so, what change in human nature, environment or culture brought about – or necessitated – religious thought and practices? Around 5 million years ago, we shared our closest relative with the chimpanzee. During the approximately 500,000 years to follow, our paths separated, producing an ‘ancestral’ chimpanzee and a distinct human ancestor: *Australopithecus*. One may well ask whether this last common ancestor practiced some form of religion. This ancestor, unfortunately, is long extinct and unable to provide insight in his religious behavior.

It is argued that the closest thing to this common ancestor we can observe is the chimpanzee: unlike humans, this species is thought to have gone through little major evolutionary changes; this is indicated by the within-species genetic differences, which are far greater for the chimpanzee than for humans. Do chimpanzees practice religion? In general, this is thought not to be the case. The closest to a ‘proto-religion’ in chimpanzees comes Goodall, who described chimpanzees displaying elaborate threat displays in the presence of rushing streams and in the midst of thunderstorms (Goodall 1975). She took this as evidence of proto-religion, assuming the behavior to be directed to either the streams and storms themselves, or to a transcendent force in an attempt to ward off its admonition. However, following a definition of religious ritual by Lawson and McCauley, Bering (2002) consistently argues that this behavior does not qualify as religion: “if chimpanzee rain dancing is to qualify as religious ritual, those enacting it must be intentionally sending a message they have encoded in symbolic action to be received and decoded by the mind of either the animated event or the transcendent force generating it; they must also represent the recipient as an intentional agent capable of knowledge acquisition”. Rather, he argues that this behavior falls under the definition of reflexive animism, i.e. the loud noises and unpredictable movements activate a hardwired behavioral response designed to scare off predators or rivals. The behavior lacks the attribution of psychological agency (higher order intentions, knowledge, belief), an essential component of religious ritual.

However, this still leaves open the question on the origin of religious thought. It has been consistently argued that this origin can be found in the ‘cultural explosion’ occurring 30,000 to 60,000 years ago (Mithen 1999). This period reveals the first material products of art, technology, and also of animistic religions (indicated by drawings of creatures that are part human, part animal). Based on this evidence of some form of religion, we can exclude the
possibility of religion having first arisen beyond this period. In addition, Mithen (1999) argues that religion requires support from external symbols (e.g. rock drawings) to be able to exist at all. This provides strong indication that religion could not have arisen long before this event of ‘cultural explosion’.

1.2 Religion as adaptation

While data from primate studies and anthropology clearly point to the period of ‘cultural explosion’ 30,000 to 60,000 years ago as the origin of religion, they alone do not suggest the reasons for emergence of religion (i.e. what change in human nature, environment or culture produced it). Its joint development with human culture may lead some to believe that religion is mainly a cultural phenomenon, but evidence from behavior genetics suggests this is not the case: religiousness – i.e. the likeliness for someone to practice religion, and the degree to which religion plays a major part in his life – has been shown to be heritable (though religious affiliation is not) (D’Onofrio, Eaves et al. 1999). This suggests that religion is at least in part genetically based.

For many biologists, admitting that a species’ trait is (partly) genetically based implies defining it as an adaptation, i.e. as a trait that survived in evolution because it increased the fitness of the individuals possessing it. Consequently, we find in biology a strong tendency to point out adaptive advantages of any genetically based trait. This tendency certainly has been prevalent when it comes to religion: many authors with a background in biology have tried to pinpoint (or, more irreverent: think up) adaptive advantages of religion. For instance, the literature study by Stein (2000) at this university focused nearly exclusively on potential adaptive advantages of religion. This study pursued the idea of authors such as Wilson and Alexander, who took religion as an adaptation for living in large groups. Religion would increase the fitness of the group as a whole because it preserves social cohesion through a set of well defined, easily shared moral norms and values (Alexander 1987; Wilson 2002). This approach thus takes morality as the principal (if not sole) feature of religion, but religion clearly is much more: aside of morality, it can deal with rituals, sacrifices, and – most centrally (Boyer 2001) – belief in elaborately defined supernatural beings. It is hard to believe (no pun intended) that a need for morality alone could induce the many other significant facets of religion. The more because religion and systems of morality do not necessarily go hand in hand: some religions do not impose systems of morality, and systems of morality do not need to build on religion; they can for instance function through hierarchical structures of authority such as governments. This leads us to believe the reasons for religion’s existence likely lie somewhere else.

Explanations of religion that focus on adaptive value rarely capture the full range of any religion’s features – if not for lack of trying (Boyer 2003)! It seems one simply cannot concoct a story of adaptiveness that comprises all diverse religious phenomena. Hence, one may well question whether it is sufficient to focus exclusively on adaptive advantages of religion, or in fact, genetically based traits in general. This is supported by Gould and Lewontin (1979), who strongly criticize the domination of the common line of ‘adaptationist’ thinking. They name five other mechanism through which genetically based traits can arise
and survive in evolution without increasing the individual’s fitness as such. Among these mechanisms is the ‘spandrel theory’, which states that particular genetically based traits can be a secondary effect (or utilization) of other adaptive characteristics. For instance, blood is red because optimization for efficient oxygen transport produced the red protein hemoglobin, not because the redness itself has any adaptive value. Anyone claiming the ‘redness’ of blood evolved because blushing is useful in sexual selection would not be taken seriously. The same line of thought can be applied to other inheritable traits, in particular when it comes to human behaviors such as religion: not all behavior needs to have direct adaptive value; rather, it can be a natural result of (interactions between) various mental systems that in turn do, or did, have adaptive value. In effect, behavior can be a by-product.

1.3 Behavior as product of a modular brain

By far, the most plausible and best-supported theories on behavioral by-products stem from the field of evolutionary psychology (EP). EP is an approach to psychology – a way of thinking about brain, mind and behavior (Cosmides and Tooby 1997). Central in this view stands the modular brain: the theory that all human brains reliably develop a standard collection of reasoning and regulatory circuits – modules – that are functionally specialized and, frequently, domain-specific. These modules are assumed to be adaptations: they have been preserved in evolution because they solved particular problems faced by our ancestors.

Evolution operates slowly, easily taking tens of thousands of years for apparent simple changes. Therefore, development of adaptive brain modules must truly have taken a long time, likely millions of years. This implies the modules must be adapted to an ancient environment, one that differed substantially from our current, industrialized world. This hypothetical ancient environment is generally referred to as the environment of evolutionary adaptation (EEA). Anthropological data indicate that humans spend by far the most of their history in hunter-gatherer societies; the first hominid appeared 5 million years ago, whereas agriculture did not develop until 10,000 years ago (taking another 5,000 to become the prevalent means of survival). Hence, the human brain is likely adapted to solve the specific problems of our hunter-gatherer ancestor, not those confronted by the modern cosmopolite Homo sapiens. Examples abound. For instance, people are still more readily conditioned to fear ancient threats like spiders and snakes than far more relevant modern threats like cars and electrical outlets. Likewise, they commonly display inordinate appetites for sugars, fats and salt when these are no longer limiting resources (Jones 1999).

This does not imply that all components of the human brain were improvised from scratch in the EEA; the long path of evolution that ultimately led to the appearance of hominids shaped important parts of the brain well before humans began to roam the earth. In general, evolution primarily builds upon traits already present (Gould and Lewontin 1979). Thus, adaptations of the human brain can to a certain extent be regarded ‘extensions’ of our predecessors’ brains (e.g. those of reptiles, mammals in general, and primates). As stated by Jones: “Human social behavior and psychology, distinctive as they are, are … recognizable as variations on primate and great ape themes” (Jones 1999).
1.4 Religion as by-product

A common theme in evolutionary psychology is that many ‘high level’ behaviors may well be non-adaptive byproducts of adaptive subconscious brain functioning. Religion might well be one of such behaviors. During the last decade, as research into the psychology of religion has (re)gained popularity (indicated by recent special issues of Psychological Inquiry, the Journal of Personality and the Journal of Psychology and Theology), various authors have articulated the suspicion that religion may in fact be just that: a non-adaptive (or even maladaptive) evolutionary by-product (Mithen 1996; Kirkpatrick 1999; Boyer 2001; Atran 2002). This appears both a viable and a plausible theory, for the following reasons: first, it is the theoretically more conservative position (Kirkpatrick 1999); instead of jumping on the adaptationist band wagon for a hunt on potential adaptive advantages, one would be well advised to first evaluate whether currently known mental systems could not (inadvertently) produce the behavior under study. In addition, by-product theories of religion typically explain the full spectrum of any religion’s features far better than the various religion-as-adaptation theories. Viewing religion as by-product certainly looks a promising approach.

In this literature study, I will attempt to present a case for the religion-as-byproduct view as proposed by Boyer, Kirkpatrick and various other authors. To clearly distinguish between low-level, adaptive brain functioning and the natural, non-adaptive religious thoughts and behavior it produces, I will first discuss various EP-based hypotheses on how the human brain was adapted in evolution to solve specific problems of the EEA, setting aside the topic of religion for the time being. The subsequent chapter argues that it is a small step – in fact, a vanishingly small one – from successful operation of the adapted brain to one that produces religious thought and behavior. This chapter also provides insight in the specific adaptive mental systems actually dealing with religion, which provide us with some clues on how religion may have developed in evolution. In the final chapter, I will treat various other, secondary characteristics of religion, and discuss the by-product approach.
## 2 On human nature

To explain the complex structure and functioning of the human brain, a study of its evolution provides a useful starting point. Building on a ‘primitive’ brain that respectively fulfilled the needs of reptiles, mammals in general and, specifically, primates, evolution likely shaped the characteristic parts of the human brain during the past 5 million years. Since *Homo* spent by far the most of that time in hunter-gatherer societies, its brain is likely adapted to the specific problems faced by our human ancestors. A careful study of the nature of these problems – respecting the history of the brain before hominids, as this provided the framework our brain was built upon – can (1) suggest the kind of structures that may have developed to solve recurrent hunter-gatherer problems, (2) suggest the design of such structures (i.e. how these problems would be solved) and (3) reveal whether a hypothetical set of adaptive structures would suffice (i.e. whether other additional adaptations are likely) (Duchaine, Cosmides et al. 2001).

This approach has certainly been taken to heart in evolutionary psychology (EP). A veritable tutti-frutti of specialized adaptive structures has been suggested to have a place in our brain. Purposes they purportedly serve include, amongst others: processing/producing phonology, musical tone and rhythm, evaluations of landscape, taxonomies of living things, individual face recognition, facial expressions of emotions, assessments of physical attractiveness and ‘mate value’ of potential mates, and theories of mental states of others. However, many of these theories on adaptive structures at best received a lukewarm reception, and frequently even met with strong criticism. Critics do not normally dispute the suggested brain functions could have had adaptive value\(^1\). Rather, they dispute whether such functions indeed are incorporated in distinct modules. The core EP states such modules are distinct (thus locatable), specialized in their domain, and will have developed (relatively) independently in evolution. However, the little evidence there is quite often does simply not support such a strong stance on the modular nature of adaptive brain functions. Hence, various respected scholars with some background in EP (e.g. Tomasello, Mithen) have chosen to acknowledge only the most basic set of brain modules.

It is most valuable to distinguish between brain *functions* with adaptive value, and distinct, specialized *modules* potentially supporting these functions. No one will dispute that the human brain can carry out all functions listed in the previous paragraph; copious evidence from psychology (as well as one’s personal experiences) supports this. Whether these functions indeed reside in distinct modules is a different matter altogether. For a consistent account of the emergence of religious thought and practices in ordinary brain functioning, the modularity of involved brain functions is, in my opinion, irrelevant. It suffices to acknowledge any human brain will reliable develop the capacity to carry out the many distinct functions: what matters is the *universality* of particular brain functions, not their potential modular nature. Even then, some may take a skeptic stance towards the ‘a small step for

\(^1\) Although it has been argued that stories on adaptiveness are always easily construed, and if proven wrong, concoction of a replacement is a trivial task (Gould and Lewontin 1979).
evidence, a giant leap in presumptions’ approach, as sometimes may loom in this chapter. In that case, I can only state that, in my opinion and in that of numerous scholars (e.g. Tooby & Cosmides, Boyer, Kirkpatrick), it sketches a plausible picture of the functioning of the human brain.

In this chapter, I will describe a number of adaptive brain functions suggested by various authors to play a role in the development of religious thought and behavior. These include the human means to detect and classify the variety of objects encountered in the environment, and to produce expectations to their likely behavior. In addition, I will devote a section to various aspects of human social behavior. It is commonly argued that the features most characteristic of the human brain deal with interaction with other individuals of our species.

2.1 Object detection

Suppose you are an early hominid, living 3-5 million years ago, roaming the savannah in small groups, hunting for prey and gathering the more sedentary types of food. What problems would you face? Let’s start with some of the obvious: you would need the ability to detect, chase and capture prey (and detect predators to avoid them chasing and capturing you), recognize food supplies, potential tools and habitats. In other words, you need to distinguish and classify the various objects encountered in your environment.

This all begins with incoming sensory information, e.g. vision (undoubtedly the most important sense in primates and humans), hearing and smell. These senses primarily present ‘raw’ information, the eyes for instance present simple two-dimensional state maps of light intensities; nowhere near the three-dimensional, full-color, categorized and moving world we experience without conscious effort. Much subconscious processing is done to arrive at the world as we know it. Many known specialized circuits are involved along the way, analyzing the shape of objects, detecting the presence of motion, judging distance or analyzing color (Cosmides and Tooby 1997). Having passed this circuitry, you observe a set of variously colored shapes, potentially moving at different speeds and in different directions. Where do we go from here? Strong evidence suggests the next step comes down to some sort of classification: are you observing an inanimate object (e.g. a stone, plant, or man-made tool) or an animate one? If the latter, is it an animal or a human? (as one would hate to mistake the approaching lion for fellow human, suitable to socialize with…) And one step beyond: if it can be identified as human, who is it? The ability to distinguish kin from non-kin, and friendly members of your community from outsiders or enemies certainly is a valuable one.

Knowing the type of object you encounter allows you to produce inferences, i.e. expectations on the object’s as-of-yet unobserved characteristics and behavior. Will it obey only the laws of physics (inanimate), or is it in addition propelled by some internal force (animate)? Was it made to serve a particular need, or could it be used to do so (inanimate objects, specifically tools)? Will it pursue some goal (animals, humans)? Is it likely to be involved in social interaction (human), and if so, what is its position in your social life (kin vs. non-kin, in-group vs. out-group)? Strong evidence suggests not only that such lines of reasoning (whose select activating come down to implicit classification) occur in humans, but
also that they are hardwired in the brain, and limited in number (Boyer 2000). This evidence is mainly based on studies of human ontogenic development: how do children perceive the world, and how does their perception change during development? It has been shown that, depending on the object presented, various and highly specific inferences are triggered in children. For instance, they may assume the structure of the object is explained by its function (an intuitive teleological stance, applied to tool-like objects), assume it has goals, eats, lives and dies (animals, humans), assume it has unobservable beliefs and intentions that drive its actions (humans). Such specialized lines of reasoning have been typed ‘inference modules’ in evolutionary psychology; these are more elaborately described in section 2.2.

Evidence for the existence of such distinct, hardwired inference-producing systems seems consistent (Boyer 2000). One of the most intriguing aspects about them, however, is their activation. What specific information triggers the various types of inferences in the human brain? One might assume the brain contains an ‘object encyclopedia’ that maintains descriptions for every type of object encountered. Incoming sensory information could then be matched against all descriptions, ultimately resulting in selection of the best corresponding description: the object is identified, and inferences can be produced. However, Boyer (2001) argues this view is misleading, if not incorrect. Rather, object detection can be viewed as a hierarchically structured set of templates (figure 2.1). As sensory information arrives, it is matches against a small number of highly abstract templates; if one such template matches, corresponding inferences are triggered, and the information is allowed to be matched against its own ‘deeper’, more specific templates. For example, only if the object has been found to be an animal, the brain is willing to consider it might also be a bird; no brain will under normal circumstances even consider the suggestion that a hammer might be a bird.

Templates typically contain only the information relevant for their complete category; a parent template should be valid for all children. Thus information in high-level, abstract templates is limited. For instance, the standard animal should describe all animals, and can therefore only detail basic features (e.g. ‘can move at its own accord’). Deeper templates on the other hand only describe aspects that are unique to their category, but lacking in the parent template. For example, the bird template could specify ‘has two wings and two legs’, and ‘can fly’\(^1\), but does not require ‘can move at its own accord’; the latter inference is already provided by the animal template.

The beauty of the concept of template hierarchies is that it allows for a clear role of learning. Children are not born with a complete template hierarchy; rather, they start out with only the highest-level, abstract templates, likely even only rudimentary versions thereof (e.g. animal is probably something like: ‘moves in ways unaccounted for by the laws of physics’). As children encounter more and more objects, they expand their hierarchies (by adding more specific categories below abstract ones), and refine existing templates (e.g. by adding more features to them). Over time, they thus both become able to distinguish more types of objects\(^1\) Clearly, this would put you in trouble when observing ostriches; when classified as birds, the conclusion they can also fly then seems unavoidable. For the sake of efficiency, the bird template likely does contain the ‘can fly’ inference; in that case, the (shared) template for land-dwelling birds should explicitly store something along the lines of: ‘cannot fly!; ignore previous inference on flying; cannot fly!’.

\(^1\)
Figure 2.1. How object perception and classification could operate: a hierarchy of templates in the brain. Initially, information from the senses is processed, revealing an unidentified object. This alone already triggers inferences on objects in general (intuitive physics). The next templates distinguish between animate objects (animals and humans) and inanimate ones. If the object is deemed animate, inferences on animate objects in general are triggered (intuitive biology), and the sensory information is passed to templates distinguishing animals from humans. If at that point the object is found human, human-specific inferences (intuitive psychology) are triggered.

(through expansion of the hierarchy), and get better at recognizing known types (through template refinement). Object recognition and classification is an always-improving skill. One might argue this implies our system for object identification is not (fully) hardwired, thus cannot be regarded a standard human feature. However, the capacity to learn – i.e. expand and improve on hierarchies – is hardwired. Thus, it seems plausible that any human in a normal environment will reliably develop the basic parts of the template hierarchy. Even some parts of the hierarchy that develop after birth can be considered universal components of the brain.

The sensory information capable of triggering an inference system may be deceptively simple; little is required to start perceiving an object as animal or human. This has been beautifully shown by Heider and Simmel (1944), who presented subjects with a screen on which a number of triangles moved according to pre-specified patterns. As the complexity of the patterns increased, subjects readily began to discern targets and motives of the individual dots, describing traingles chasing and avoiding each other, and forming social groups (of like-minded individuals). Thus, particular patterns of motion sufficed to perceive simple triangles as agents. A similar example can be found closer to home: taking a walk through a forest, you suddenly hear the noise of a branch falling down a tree. The primary, initially subconscious response is to presume to presence of an agent, be it animal or human. Only much later – in relative terms –, taking into account the strong gusts of wind experienced before, you might reluctantly drop the suspicion of agency, and attribute the event to the weather. Agency
detection clearly does not need too elaborate stimuli, and could hardly be called adaptive if it did; the lion would have consumed you, long before you realized the four legs, tail, fur, and many ferocious teeth moving at such high speeds in your direction could potentially present a danger to your health.

2.2 Inference modules

Based on information received from the senses, humans classify the objects in their surroundings. Here, classification means little more than the activation of one or more ‘inference modules’: modules incorporating particular ways of thinking about objects. Evidence from various fields of research (e.g. brain section pathology, cross-cultural consistencies), in particular that of ontogenic development, indicates the human brain contains of limited number of such modules. Most commonly mentioned are modules for ‘intuitive physics’ (thinking about how the laws of physics affect objects), ‘intuitive biology’ (thinking about the properties of the living, particularly animals and humans) and ‘intuitive psychology’ (thinking about an object – typically a human – as one with a private set of beliefs on the world around it and a set of intentions) The next section describes various aspects characteristic of these modules.

2.2.1 Intuitive physics

Any object, non-living, animal or human, is subject to the laws of physics. It would certainly be adaptive to realize this as you observe (the motion of) objects, and ponder on interacting with them; any agent lacking even the most basic feel for physics would all to ready walk off a cliff due to ignorance of gravity, or crash into the nearest object in its way because he failed to take into account its solidity… This is not to say that agents should be adapted to subconsciously memorize Newton’s laws and other essential definitions in physics; this would be far too bold a statement (let alone that it would strongly reduce the job availability for physics teachers). However, a basic, ‘intuitive’ feel for physics would certainly prove useful, if not indispensable.

Copious evidence indicates humans possess an inference module that deals with intuitive physics. Very young infants already appear to grasp certain key physics-related principles, such as ‘continuity’ (objects move in continuous paths), ‘solidity’ (objects do not coincide in space), ‘support’ (unsupported objects fall downward) (Baillargeon 1987; Baillargeon and Hanko-Summers 1990; Spelke 1990) and the rule ‘no action at a distance’ that excludes non-contiguous physical causation (Leslie 1988). This has been demonstrated in various experiments; infants, for instance, seem surprised when confronted with two solid objects seemingly passing through each other. The basic principles of physics are applied to any object encountered (artifact, animal, human). However, their dominance in the train of thought about objects under study is dependent on activation of other inference modules (biology, psychology), which tend to overshadow the inferences on physics.
2.2.2 Intuitive biology

Back to the proverbial lion. A grasp of the basic principles of physics no doubt is highly practical in dealing with objects encountered. However, it would leave one wanting if he allowed it to fully dictate his anticipations on lion behavior. Valid rules for inanimate objects (e.g. non-contiguous physical causation is an impossibility) simply are insufficient when it comes to anticipating the behavior of living things. Humans spending a significant amount of their time hunting prey (and avoiding predators) would certainly benefit from a good understanding of the behavior of living things.

Evidence indicates humans possess a specialized inference module that allows for reasoning about the characteristics and behavior of living things. First, and perhaps the most important, this module supports the presumption that living objects are propelled by some internal force (as opposed to only physical interactions). Children already grasp this principle: to preschoolers, it is quite clear real animals can move at their own accord, while statues of animals cannot\(^1\) (Massey and Gelman 1988). In addition, the inner propelling force is assumed to pursue specific goals: our intuitive biology supposes the directions and patterns of movement are typically dictated by specific goals inherent to the living (e.g. search for food, avoidance of predators, pursuit of prey). Intuitively, when humans observe a living object move, they evaluate its potential motives: why is that cat sneaking up on the bird? – Former intends to catch and consume latter. Why is that rabbit speeding away from us? – It regards us a predator and declines to be on the menu. Such subconscious lines of questioning are certainly adaptive, as they allow us to predict future behavior and trajectories, indeed indispensable for the predator (and potential prey) we are – or at least our ancestors were.

Second, our intuitive biology appears strongly adapted to classify the various living things encountered: humans tend to think in terms of species. Various individuals may differ somewhat (or more: male lion vs. lioness) in outward appearance, but are easily mapped to one species of which the internals are always similar\(^2\). This may seem obvious, but it is not. People readily assume the internals of any tiger will be similar to the first one dissected, yet are much less sure about the internals of clocks and televisions, for instance (Boyer 2001).
The ‘species’ concept found in intuitive biology seems to be represented as a particular ‘essence’, invisible but intrinsic to every individual of the species. Boyer (2001) gives an example for cows: “Suppose you take a cow, surgically remove the excess body mass and remodel it to look like a horse, add a mane and a nice tail and perform other operations so it eats, moves and generally behaves like a horse. Is it a horse? For most people, including most children, it is not. It is a disguised cow, a horsy cow to be sure, perhaps a cross-cultural cow, but in essence still a cow. There is something about a cow that is internal and permanent. You

---

1 Preschoolers are not quite clear on the actual reasons for this difference, however. When asked why the statue could not move at its own accord, they explained that it could not, because ‘it had no legs’. When shown the statue did have legs, they argued that ‘these were not good legs’, etc. This is not truly surprising, since assumptions such as a ‘propelling inner force’ are produced by our intuitive biology; a system separated from conscious access (Boyer 2001).
2 The human brain mapping variable, imperfect individuals in the environment to abstract, ideal species: Plato would be proud!
can have this assumption without having any representation of what the ‘essence’ is. That is, most people represent that cows have some essential ‘cowness’ about them, even if they cannot describe cowness. All they know is that cowness, whatever it is, cannot be removed and creates external features. Cowness explains why a cow grows horns and hooves.” In the literature, numerous beautiful examples of this assumption of irremovable essence can be found. For instance, in one experiment children were told doctors took a raccoon, spray-painted it black with a white stripe down its back and implanted it with some ‘super smelly yucky stuff’. The children were then shown the picture of a skunk, and asked what it was. Most of them stated that it was a raccoon. On the other hand, when told doctors took a coffee pot, sawed off its handle, cut a hole through it and filled it with birdseed, and then are shown the picture of a bird feeder, they consistently identify the object as a bird feeder (Pinker 1999). Clearly, coffee pots are not assumed to possess an irremovable essence. A similar experiment was done with a tiger made to look like a lion, and a wastebasket used as a chair; this produced the same result. Irremovable essences are assumed specific to the living.

Humans assign every living individual encountered to a species with its intrinsic essence, and one species only. In addition, they appear to construct a taxonomy of species: each species is placed in a nested hierarchy of kinds, with no overlap between species at the same level (Atran 1990). This could be considered an intuitive, Linnaean type of classification\(^1\), with a number of encompassing categories (mammals, birds, reptiles, etc.), and deeper, fine-grained distinctions. This type of classification seems to be exclusively applied to the living world; it does not represent a general mania for classification because it is in this form (i.e. with strict rules: every individual belong to a species, classes cannot overlap) rarely found in classifications of non-living things (Atran 1990). For instance, a piano can be both furniture and a musical instrument; it thus belongs to two classes simultaneously.

Why would strict reasoning in terms of species and encompassing hierarchies be adaptive? Primarily, it allows for generalization: various encountered individuals are mapped to one single species; characteristic features (e.g. appearance, but also predator/prey behavior and edibility) of these individuals need therefore to be stored in memory only once: as belonging to the abstract species. In addition, the intuitive ontology of living things allows one to predict properties of species never encountered before; based on outward appearance, the individual can likely be assigned to one of the more general categories (e.g. mammals, birds). This automatically defines some of the hidden features the individual is likely to possess: this is some sort of bird; if I run very fast to capture it, it might be inclined to fly away. Classification of living things certainly would prove useful.

Various experiments have indicated the human intuitive biology produces a number of other inferences. For instance, children assume living objects eat, produce additional individuals of the same species and die (Boyer 2001). Not only that, they seem to make elaborate assumptions on such subjects. For instance, children’s understanding of death in

\(^1\) It is likely no coincidence that scientifically accepted classifications of species tend to be hierarchically structured. Undoubtedly, the human disposition to think in hierarchical structures enabled both the classification theories themselves, and their acceptance.
intuitive biology context appears well developed; when shown film clips of successful
predation, children readily understand that (1) the (killed) prey is no longer conscious, i.e.
does not have goals, (2) the individual no longer contains an ‘inner force’ that produces
animation and (3) this situation is irreversible (Boyer 2001). Such biology-based inferences
clearly could prove adaptive far the hunter-gatherer hominid.

2.2.3 Intuitive psychology

Anthropological evidence indicates early hominids lived in groups, in essence small social
communities. This does seem the more probable because most primates and monkeys
(suitable templates for our last pre-\textit{Homo} ancestors) also form close communities. To enable
successful living in groups, however, one needs understanding of the intentions and
(potential) actions of others: how will individual A respond if I do this? What do that actions
of A imply, what does A desire? One needs a means to understand and interact with fellow
members of one’s species. Evidence from ontogenic development and brain pathology
indicates humans (and to a lesser extent, primates also) possess an ‘intuitive psychology’
inference module, specialized in reasoning about other individuals of the species.

Inferences produced by our intuitive biology already comprise many of the features
applicable to humans: they are propelled by some inner force, their patterns and direction of
movement are governed by their goals, they eat, reproduce and die. What more could one
need to live in human social groups? Intuitive biology does provide the assumption that living
things have goals; however, the set of goals it can attribute is limited, as it mostly restricted to
primary biological needs (e.g. need for food, avoiding predation). Also, our intuitive biology
cannot attribute goals based on the static image of individuals; rather it derives them from the
course of action they follow. This in addition implies it does not support a per-individual
memory of the current state of mind, and cannot instantaneously anticipate on the actions
individuals will perform. This various shortcomings would make our intuitive biology alone a
very primitive, inefficient and lacking means for close interactions with fellow humans
indeed! More is needed to attain adequate and elaborate understanding of humans. This need
is fulfilled by our intuitive psychology module.

The key assumption provided by our intuitive psychology is that the behavior of persons
is caused by unobservable beliefs and intentions. This essentially comprises three
components: every individual (1) takes his own view on the world, dictated by his
observations and experiences, (2) houses various intentions and (3) decides on a course of
action based on his current world view and intentions. Thus, a per-individual memory of his
‘world view’ and intentions would allow one to anticipate on behavior of others, both in real
(i.e. current) and hypothetical scenarios. This is precisely the capacity our intuitive
psychology provides.

We are capable of understanding others are partly driven by their beliefs on the world
around them. Here, the title ‘belief’ is most relevant: our intuitive psychology appreciates they
are only just that: beliefs (i.e. minds \textit{think}, rather than \textit{know}). This implies that between
individuals, beliefs can differ and conflict, as humans have imperfect access to information.
Simply put, one’s set of beliefs is not all compassing, and some beliefs can be false.
Only humans beyond the age of four appreciate others may harbor false beliefs; various developments in our intuitive psychology precede this assumption of false belief. At the age of 6 months, infants only engage in dyadic interaction; they take only themselves and the adult into account simultaneously, and thus are limited to one-on-one interaction. Between the age of 9 to 12 months, infants engage in triadic interaction; they become capable of grasping scenarios involving themselves, an adult and an outside entity: they can follow the gaze of the adult (What object is he looking at?) and perform imitative learning (repeating actions involving the outside entity, actions as seen performed by the adult). According to Tomasello, this is the moment children start to perceive persons as intentional agents: animate beings with the power to control their spontaneous behavior, and in addition harboring goals and making active decisions among behavioral means for attaining these goals (Tomasello 1999b; Tomasello 1999a). Note that this already is more complex than intuitive biology thinking, because the agent in question makes active choices among behavior means. From the age of three, children assume that actual states of affairs cause perceptions that cause beliefs that cause intentions; they also understand that these beliefs and intentions are immaterial (Boyer 2000). Then around the age of four enters the assumption others beliefs may well be false. This has been elegantly proven with so-called ‘false belief tests’. One such test, described by Fikes (2001), runs as follows: the child is placed in front of a stage, and asked to watch what takes place. Enter Sally, who has a marble that she places in a basket and covers, then departs. While she is gone, Ann removes the marble from the basket and places it into the box. The child’s task is then to predict where Sally will look when she returns. Children up to the age of three answer Sally will look in the box (‘We all know the marble is there, right?’); from the age of four they can appreciate Sally has not witnessed Ann relocating the marble, and will therefore believe it to be still in the basket.

To what extent can we regard the intuitive psychology inference module a human-specific feature? Jones describes two examples that indicate chimpanzees also possess some (more rudimentary) form of intuitive psychology. First, when shown film clips of human experimenters trying to solve particular problems, chimpanzees can point out the tools best suited to solve these problems. Second, they can recognize that the experimenter who accidentally deprived them of food last time is more likely to help than the one who seemingly acted with deliberate malice (Jones 1999). In addition, many accounts of monkeys (i.e. species older – more ‘primitive’ – than the chimpanzee) in the wild suggest that in some cases, A can take into account that what B sees is different from what A sees, and can use this knowledge in tactical deception. These examples indicate chimpanzees and monkeys may be capable of representing the knowledge and intentions of others. Lacking, however, is any indication of appreciation of ‘false belief’; this does appear to be a human-specific feature (Jones 1999).

It is worth noting that several authors take a more skeptical stance regarding the presence of a primitive intuitive psychology in chimpanzees (Tomasello 1999b). Generally accepted, however, is that the groundwork for human intuitive psychology was already laid in primates. Opinion differs mostly on specific capabilities supported by this primitive version.
Much of the evidence for an intuitive psychology module stems from studies on autism. Autism is in essence a deceptively simple diagnosis, as it seems to comprise a number of different afflictions, which can differ considerably in severity over patients. However, most of its manifestations are typically associated with (partial) failure of the patient’s intuitive psychology module (Baron-Cohen, Leslie et al. 1985; Baron-Cohen 1995). Autism patients appear unable to appreciate other harbor private, potentially false sets of beliefs. This has again been well demonstrated with false-belief tests: autism patients typically fail these. Much information – particularly regarding modularity and localization – on our intuitive psychology can be obtained from autism studies.

2.3 Social systems: kin, friends and communities

Humans today live in social communities, as did the first hominids 5 million years ago. In fact, many, many species live in groups: from the hymenoptera order of insects (bees, ants, wasps) up to numerous species of mammals (to name a few: ground squirrels, bats, horses, wolves and all ape species except the orangutan). But why would any species form social groups? What adaptive advantage could a social structure bring? Certainly, if every individual group member went its own way, ignoring others, group formation could only be qualified maladaptive: it would just infuse resource competition and disease spreading. Obviously, groups do not function this way. Group members help each other, either directly (e.g. through sharing food and other resources), or by taking on specific tasks best suited to their capabilities (for instance, young mate-less male chimpanzees practically qualify as group soldiers, marching at the outside ring of a moving chimpanzee group, and being the first to engage in battle). Groups revolve around altruistic behavior: helpful interaction between group members.

Altruistic behavior originally placed evolutionary biology for a dilemma: since we generally accept those individuals with the highest fitness (e.g. spending the least amount of effort to acquire the largest amount of resources, mates, etc.) will benefit (i.e. overrule) in evolution, why would any individual engage in costly group aid? Suppose the trait ‘helping others’ emerged in evolution, and even got as far as spreading across an entire group. The first group member losing that trait (which should not be too difficult; as in most contexts, in evolution it is easier to demolish than to build) would certainly possess the greatest fitness; every member would support him, without him experiencing any ‘foreign aid’ costs.

2.3.1 Parental care

Let’s start with on of the most basic forms of social interaction: interaction between parent and child. Parents tend to feed, protect and generally care for their offspring. This comes at great costs; they forsake food they could have eaten, might be hurt – even killed – in defense of their young, and invest great amounts of energy in the latter’s well-being. Costly parental care occurs in humans, as well as in many birds and other mammals. Yet, this does not pose a problem for evolutionary biology, as production of successful offspring (i.e. healthy offspring reaching reproductive age) is a defining part of one’s fitness; one cannot spread his genes simply by eating most, growing fastest, and living longest.
Parental care certainly is adaptive. But which mechanisms in the brain support the typical parent-child interactions? To start with, parents need a mechanism that allows them to recognize their offspring, be it through vision, smell, hearing, or any other sense. In addition, sophisticated parental care (such as that of humans) requires the parent to experience some sort of empathy: they must deduct the emotional state of their offspring, and attempt to alleviate negative feelings, and enhance positive ones. Evidence indicates we likely possess specific systems that allow us to feel empathy. First, the human brain appears to have a hardwired capability for expressing emotion; cross-culturally, one finds the same stereotype facial expressions for six universal emotions, termed happiness/approval, rage/anger, fear, grief/sadness, disgust and surprise (Ekman 1998). Correspondingly, humans also possess a specialized system capable of deducing emotional states of others, specifically from facial expressions (Gray, Young et al. 1997; Broks, Young et al. 1998; Adolphs, Tranel et al. 1999; Blair and Curran 1999). The ultimate feeling of empathy itself would be an analog of the means utilized to analyze movements of others: our brain simulates witnessed movements of others, as if it were performing them itself. These simulations are elaborate: they utilize the same pathways as own motor actions, and are blocked only at the point actual motor neurons would be fired (this blockage does not yet occur in young infants, resulting in imitative behavior). Several authors have suggested that we may possess an analog of this system that simulates emotions of others (Boyer 2001; Hurlbut 2001); if effect, we would be feeling their (witnessed) pain if it were our own. As with our own pain, we would be inclined to alleviate theirs. Of course, such a system for empathy would require select activation; it would have to respond most strongly – perhaps solely – if observing our own offspring.

Systems for offspring recognition and empathy support the basis of parental care, but do not illuminate the specifics of child→parent interaction. In many species characterized by an extended period of development and parental care, mechanisms are activated in offspring to capitalize on and maximize the receipt of parental investment (Kirkpatrick 1999). Such mechanisms would include a means of recognizing the parents (naturally), as well as an ‘attachment system’ that is activated by clues of potential danger (causing the child to seek safety with the parent) and allows the parent to be viewed as a safe haven for confident exploration of the environment (Kirkpatrick 1999). Many authors have maintained that this ‘attachment system’ also plays a prominent role in adult love relationships (substituting the partner for the parent).

Parental care – even with associated extreme costs – is unquestionably adaptive, and may therefore seem uninteresting from an evolution-of-altruism point of view. However, it is important to realize that the existence of parental care implies an infrastructure for aiding and caring about others is already in place. This ‘parental care’ infrastructure has been thought to provide the framework for other social behavior (Kirkpatrick 1999): social behavior towards non-offspring might (partially) operate by extending child recognition systems to include other individuals.
Figure 2.2. Mechanisms for parental care. As the parent observes and identifies offspring, its systems for deduction of emotional states are stimulated. Subsequently, the offspring’s deduced emotions are ‘simulated’, resulting in empathy. This induces altruistic behavior. At the same time, the sight of the parent could cause offspring to enhance expression of emotions. Note that the attachment systems that determine the behavior of offspring are omitted; these are not involved in parental care.

2.3.2 Kin selection

In addition to clearly adaptive caring for offspring, various species also display altruistic behavior towards other kin members (e.g. siblings). For instance, in many bird species, flown-out young assist parents in caring for subsequent nests, rather than starting their own. Thus
they aid their parents (and siblings) at the expense of their own reproduction¹ (Gadagkar 1997). Like parental care, this seems still altruism in the true sense of the word: aiding others at the expense of oneself, without requiring anything in return. Now what could be the adaptive advantage of that?

Fitness in essence comes down to the ability to spread one’s genes. As a result, any means of spreading one’s genes is adaptive. Now kin members share part of one’s genetic material (e.g. for diploid species, siblings share 50%, cousins 12.5%). Increasing the fitness of kin members thus indirectly enhances the propagation of one’s own genes. Such a fitness increase of kin may even come at the cost of decreasing one’s own fitness, provided his inclusive fitness – the expected overall propagation of one’s genes, both in oneself and in kin members – increases. To illustrate the concept of ‘inclusive fitness’: if you experience a decrease in fitness of 50%, while your sibling as a result thereof experiences a 100% increase, the net change in inclusive fitness is zero. This led to the interesting – and hopefully purely hypothetical – example that you could self-sacrifice in the act of saving 9 drowning cousins (with the premise their fitness was equal to yours to begin with); this would produce a 12.5% increase in your inclusive fitness.

The mechanisms that produce kin selection in humans have been little studied. However, kin selection behavior shares many features with typical parent→child interaction; it requires one to recognize individuals (i.e. kin), deduct their emotional state, and trigger feelings of empathy. It may therefore well utilize the systems involved in parental care. However, it clearly does require one additional feature: it needs to allow for gradation of altruistic behavior. Depending on the kin member encountered it should trigger more (e.g. toward siblings or less (e.g. towards cousins) altruistic behavior (and empathy). And this behavior should never become as costly as that directed towards own offspring. Self-sacrifice for the son of your grandmother’s cousin (sharing 1.6% of your genes) would not increase your inclusive fitness.

¹ It is worth noting that a number of other hypotheses might explain this behavior without involving kin selection. These come down to the following: (1) helpers have a better chance of survival if they associate themselves with a nest; (2) by helping, helpers increase the probability of becoming breeders in the future; (3) helpers are more successful at breeding if they have had past experience as helpers. However, at least in some cases (e.g. the white-fronted bee-eater), there is clear evidence against these hypotheses; in such cases, only kin selection can explain the helpful behavior (Gadagkar 1997).
Figure 2.3. Mechanisms for kin selection. As an individual observes and identifies kin, its systems that deduce emotional states are stimulated. Deduced emotions are subsequently ‘simulated’, resulting in empathy. This then results in altruistic behavior. At the same time, the sight of kin could induce the other individual to enhance expression of emotions.

Note that both individual are equal: they both possess the full set of systems. This picture simply distinguishes between the favor-providing altruist and the recipient, and displays only the systems relevant for their particular roles.

The systems for kin selection are very similar to those involved in parental care. In fact, only one small thing has changed: the route from individual recognition to empathy (and expression of emotion) has been expanded to include individuals aside of offspring or parents. However, the sight of kin is not likely to produce such a vehement response as sight of offspring; that would induce a maladaptive level of kin-directed altruism. As more paths came to lead to altruism, the strength of the altruistic response likely became better controllable.
2.3.3 Reciprocal altruism

Not all human social interactions involve kin. Our altruistic behavior goes beyond caring for offspring and kin selection. For instance, humans involve in reciprocal altruism, which comes down to providing aid to other individuals that they will (in time) reciprocate. Such behavior is not limited to humans. For instance, vampire bats that successfully fed on farm animals are known to share blood with individuals that failed (Gadagkar 1997). This may seem true altruism, but in effect is more of a ‘tit-for-tat’ strategy: the sharing is reciprocated. If A shares blood with B, B will share with A on future occasions; if B defects, A will no longer share with B.

For reciprocal altruism to be adaptive, certain conditions must be met. First, individuals failing to act reciprocally must be excluded from further interaction (i.e. sharing). Given enough time, there will undoubtedly appear an individual in a reciprocal society that does not reciprocate. If other individuals continued sharing with that particular individual (which comes down to increasing its fitness at the expense of their own), its fitness would experience an inordinate increase; it would be able to produce more, and more successful offspring (more successful as part of its offspring likely shares its tendency to not act reciprocally). Over time, reciprocal individuals would all be replaced by non-reciprocal ones. Non-sharing individuals must thus be excluded from further interaction. Primarily, this requires one to detect these ‘cheaters’. This calls for a memory of individuals aided, and of those that aided oneself. This may or may not go combined with a memory of the ‘value’ of the favor provided; instead of memorizing that you owe B, or B owes you, you might elaborate on that by remembering the worth of the favor owed. The second condition for reciprocal altruism is that the individuals involved maintain a stable, long-term relationship: if one aids B, but B leaves the next morning never to be heard from again, the chance of reciprocity is non-existent.

Clearly, humans involve in reciprocal altruism. For instance, truly small communities (e.g. isolated villages) could be considered functioning through reciprocal altruism; typically, one finds in such communities that one knows other community members quite well, and is willing to aid. A per individual memory of reciprocal states (owing this and that much) then is feasible. Also, one could perhaps qualify (long-term) friendship as such: friends help each other, support each other in times of need, in the expectation will help and support them in similar circumstances. One would hardly term someone consuming your aid and support, but never reciprocating it, a friend (rather, a nuisance – or a client, and charge him, if you happen to be a psychologist). What mechanisms support reciprocal altruism in humans? Humans already possess a sophisticated architecture for recognizing individuals; this already serves us in identifying offspring and other members of our kin. This could provide the framework for a per-individual memory of reciprocal states. Evidence also indicates humans have a well-developed mechanism for social exchange reasoning (Kirkpatrick 1999). This would support sophisticated ‘tit-for-tat’ thinking, by determining the worth of the favor given or received, and the type of favor (i.e. its worth) acceptable to provide in return. In addition (or perhaps as part thereof) to social exchange reasoning, humans possess a basic mechanism for cheater detection, which identifies individuals that fail to act reciprocally (either through altogether
failure, or ‘sneaky failure’: not returning favors of the same ‘worth’); such individuals can then be excluded from further favors, potentially even punished.

Kirkpatrick mentions an interesting side effect of our mental architecture for reciprocal altruism. For anyone to engage in this form of behavior in the first place, one must believe favors are typically (in time) reciprocated. This might be the cause of the highly common belief in a just world, where ‘people get what they deserve’ (Kirkpatrick 1999).

Figure 2.4. Mechanisms for reciprocal altruism. In this picture, individual A (the potential altruist) owes individual B (the potential recipient). Both individuals know each other, thus maintain the other party’s reciprocal state in memory (i.e. in worth favors owed by the other, or to the other).

As A identifies B, the system for social exchange reasoning recovers the state of their reciprocal relationship. This should induce A to act altruistically towards B. On the other hand, B identifies A, recovers (his version of) their reciprocal relationship, and expects A to act altruistically. To ensure this, B’s system for cheater detection monitors the behavior of A, and will stimulate positive behavior (i.e. rewarding) if it witnesses altruistic acts, and negative behavior (punishing) if A fails to act altruistically.
2.3.4 **Indirect reciprocal altruism: communities**

Very small, stable social communities might function through reciprocal altruism only – with their members experiencing a joint fitness increase. But what happens when the population becomes larger? Clearly, there are limits to an individual-based memory as needed in reciprocal altruism; in a city like New York, keeping records for every member of the population – in excess of 10 million – would prove a daunting task indeed for the brain! Combined with typically larger migration rates in such communities, reciprocal altruism would undeniably prove a maladaptive strategy if applied to all individuals encountered. Yet, altruistic behavior does occur in large, communities, albeit often in reduced form (i.e. smaller favors are provided, less cost is inflicted on one’s own fitness). Why is that?

Alexander argues our hunter-gatherer ancestors could experience a number of advantages from group formation: aggregation at common resources (which then permits information sharing), predator avoidance, active defense against predators, nepotistic investment in kin, and cooperative hunting and killing of otherwise unattainable large prey. These advantages can explain the emergence of small, partially kin-based groups, but fail to explain the large size of human societies, and their growth in size in history (Roes and Raymond 2003). In fact, “as hunting weapons and skills improved, group sizes should have decreased” (Alexander 1987). A commonly suggested solution to this paradox is that other human groups were the ‘predators’ that prompted people to live in ever-larger societies; inter-group competition fueled increases in group size. Inter-group competition in humans is not unlikely; we likely inherited the tendency to raid and compete with other human groups from our last common ancestor (in common with the chimpanzee), as chimpanzees act similarly: they are known to form multi-male coalitions to raid neighboring groups (Duchaine, Cosmides et al. 2001). Systematic group-based aggression towards members of one’s own species likely first developed in this common ancestor: “only two species are known … [to employ] a system of intense, male-initiated territorial aggression, including lethal raiding into neighboring communities in search of vulnerable enemies to attack and kill. Out of four thousand mammals …, this suite of behaviors is known only among chimpanzees and humans” (Wrangham and Peterson 1996).

The formation of large human groups is likely to have been adaptive, from the groups’ point of view. However, to stay out of the dangerous waters of – much criticized – group selection theories (stating that adaptive traits can be useful for the group as a whole, but need not to be for individuals themselves), we can add one premise: in evolution, viability of the tendency to form large groups requires a means of excluding uncooperative individuals (i.e. cheaters). Only then is formation of large groups of altruistic members adaptive for the group as a whole, as well as for individuals participating.

Let’s take a closer look to the characteristics of indirect reciprocal altruism. As we saw before, large groups require a shift from a per-individual memory (of reciprocal states) and

---

1 The bonobo (the chimpanzee’s closest cousin) does not involve in inter-group competition (de Waal 2001). However, this does not imply that the same applies to our last common ancestor (of chimpanzees, bonobo’s and humans); the bonobo may have lost the trait for inter-group competition in evolution.
associated cheater detection to a system that allows for altruistic interaction with anonymous individuals. In addition, this system should function similarly in all group members, i.e. they should all perform altruistic actions of the same worth in particular situations. This comes down to morality: all individuals (in a certain group) share identical norms and values that determine how one should treat other, potentially anonymous individuals.

What mental systems are needed to support morality? The principal component – again – is thought to be an extension of mechanisms for parental care (in particular: empathy). Instead of applying these to offspring only, the system could have evolved to include other individuals. This does not imply it should necessarily serve any and all other individuals. Recall the increase in group size is likely due to inter-group competition. In this context, altruistic behavior should primarily be directed towards members of one’s own group. Rather than bypassing individual recognition altogether, the system’s focus should shift from ‘which particular individual?’ to ‘does the individual belong to my group?’. Evidence indicates the human brain drafts its intuitive biology (section 2.2.2) to this purpose: distinct groups (and in particular one’s own) are represented as distinct species, with distinct ‘natural essences’ (Kirkpatrick 1999). Individuals could recognize members of their ‘species’ through appearance (e.g. race), or language (as illustrated in Judges 12:4-6: Jephthah used the test word ‘shibboleth’ to distinguish the fleeing Ephraimites from his own men the Gileadites; the former could not pronounce the ‘sh’). As our intuitive biology allocates every species a position in the all-compassing hierarchy of kinds, this would also explain the human feeling of belonging to multiple groups (e.g. community, nation, all of humanity), with varying ‘strengths’ of membership (and varying levels of altruistic behavior). This representation of human groups as non-overlapping species could also account for typical human ‘in-group vs. out-group’ reasoning, as commonly found in sociological research (Kirkpatrick 1999).

Parental care mechanisms may have expanded to produce altruistic behavior towards others, and intuitive biology may well aid in restricting this set of ‘others’ to members of one’s own group. However, as the number of individuals one engages with in altruistic interaction increases – i.e. more indiscriminately includes new individuals in such interactions –, one can no longer expect all members to perform altruistic actions of the same worth in identical circumstances; variation between individuals would simply be too great. To resolve this problem, selection pressure for larger groups is thought to have produced well-developed capabilities for definition and absorption of moral codes (Turner 1997). These capabilities would enable all individuals to display similar altruistic behavior under similar circumstances.

What if one individual decides to break the shared moral codes? If he forsook all altruistic behavior but happily accepted favors from others, he would obtain the highest fitness, and ultimately (through offspring) replace altruistic individuals. To prevent this, a mechanism for cheater detection is required; unlike cheater detection in true reciprocal altruism, however, this mechanism now needs to activate when any group member breaks the moral codes (as opposed to only when a specific individual fails to reciprocate one’s own favors). Evidence suggests humans possess sophisticated means to enable this type of ‘moral cheater’ detection. Cosmides (1989) found for instance that humans are far better capable of solving problems in logic (detecting violations of the rule ‘if p, than q’) when these are
phrased as problems of cheater detection (e.g. a rule such as: ‘only people with tattooed faces are allowed to eat cassava’). The mind seems well designed for detection of violations of moral codes.

If moral cheaters are detected, they should be sanctioned (e.g. through exclusion from further altruistic acts, or punishment) to prevent them from achieving inordinate fitness. Preferably, mild sanctioning should be even done by individuals that were not a direct victim of the violation. This type of behavior is typical in human communities: for instance, one feels anger when observing someone cutting in line, even the observer was no part of the line himself. And beyond: humans even feel a tendency to punish those that failed to punish cheaters (Boyer 2001).

Summarizing: in indirect reciprocal altruism, as typical in large societies, most mechanisms involved in other types of social behavior resurface: they include expression and deduction of emotional states and feeling empathy (from parent→child interaction), cheater detection (which now focuses on people breaking moral codes) and sanctioning (which now may be carried out by non-victims). The most significant change that truly enabled living in large groups is the emergence of systems for definition and absorption of moral codes.
Figure 2.5. Mechanisms for indirect reciprocal altruism. Both individuals are members of the same group, thus share the same moral codes. Here, these codes specify that the situation calls for A (the potential altruist) to act altruistically towards another group member. B can be either the direct beneficiary, or a bystander that only observes the situation.

A can either come to altruistic acts through systems for empathy, or through its moral codes. The route to empathy has expanded to include group members in addition to kin. In addition, the sight of group members now triggers the system maintaining shared moral codes; if these codes specify the situation calls for altruism, A should act altruistically.

Individual B identifies A as a member of its group, which triggers the moral codes (and also, enhances expression of emotion). Based on the moral codes, B decides whether the situation calls for altruistic action of A. If so, B’s system for cheater detection monitors the behavior of A, and will induce positive behavior (i.e. rewarding) if A complies, and negative behavior (i.e. punishing) if A does not.
3 The naturalness of religion

Five million years of hunting and gathering produced a species with sophisticated means for perceiving, classifying and reasoning about its environment, and elaborate forms of social behavior. These capabilities are made possible by specialized mental systems. Such systems include modules that detect key features of objects (e.g. shape, patterns of motion) and inference modules that produce specific expectations: intuitive physics, intuitive biology and intuitive psychology. In addition, they include systems that support social interaction: means for expression and deduction of emotional states, feeling empathy, social exchange reasoning (‘tit-for-tat’), cheater detection and definition and absorption of moral codes. But why would such a species engage in religious thoughts and practices?

3.1 Pathways to supernatural concepts

3.1.1 Over-detection of agency

Specific incoming sensory information triggers inference modules (intuitive physics, intuitive biology, intuitive psychology). These modules are specifically suited to reason about particular classes of objects: our intuitive physics describes the properties of inanimate objects, the combination of intuitive physics and biology describes the properties of animals, and the combination of all three – physics, biology, psychology – describes the properties of humans. Thus, the type of object presented should determine which inference modules are triggered: if you see a stone, this should trigger intuitive physics only, but if you see a human, this should trigger intuitive physics, biology and psychology. However, evidence indicates the mechanisms capable of triggering inference modules are not that selective. In section 2.1 we already saw that simple on-screen dots can trigger our intuitive biology – perhaps even psychology – if these dots move according to specific patterns. Similar examples are described by Guthrie, who argues we in fact perceive humans all around: we see human faces in clouds and pictures of Mars, observe human shapes in oddly shaped trees, and hear human activity when branches stroke the window of our isolated cabin (Guthrie 1993). Little seems needed to trigger specific inferences.

We perceive the most complex type of object (i.e. with most inferences), a human agent, everywhere. This does not happen because it is reassuring – occupying an isolated cabin, the notion of unseen human agents seem little reassuring compared to the notion of branches stroking the window (Guthrie 1993; Boyer 2001). Rather, we perceive human agents all around because it is the adaptive thing to do. A hunter-gatherer life style involves many split-second decisions based on incoming sensory information. One must respond to shapes looming in the dark, the sound of something coming up fast from behind, a rushing of the bushes. If given the choice between perception of inanimate objects, harmless animals, predators or (hostile) humans, one would better start with the most life threatening scenario: better safe than sorry. As Kirkpatrick (1999) argues: it is better to mistake a stick for a snake, than the other way around. One can safely mistake a stick for a snake dozens of times, but the one time you mistake a snake for a stick could prove fatal. In other words, given limited
sensory information, it is adaptive to (initially) overestimate the type of object involved. In the long human history of inter-group competition and aggression, a tendency to perceive the most complex – and potentially dangerous – type of object, a human, will have proven truly adaptive. As a consequence, however, you will quite often perceive human agents when there are none: the brain creates its own supernatural agents, combining essentially a non-human object with typically human features. For instance, in South-America, a volcano is thought to offer protection if well fed: every now and again, food must be dropped into the volcano’s mound (Boyer 2001). And in parts of Africa, certain tribes maintain that ebony trees can overhear conversations of the people sitting underneath them (Boyer 2001).

3.1.2 Misfortune and social intelligence

Thus far, I have described agency detection as if it attends solely to direct mapping of sensory information to particular classes of objects. However, it does not only work to produce such direct links; it can attribute human agency through more indirect paths. Recall the ‘walk through the woods’ example of chapter two, in which you hear a branch breaking and falling down a tree. Your first impulse may be to take the branch itself for an agent, animal or human. But even if you went to investigate and found it was just a branch, you automatically look up the tree and expect to find the animal or human breaking it in the first place. Humans are predisposed to look – and keep looking – for causes that, for the sake of safety, should primarily thought to be human.

The human search for causes is difficult to satisfy. Back to the walk in the woods: even if you got as far as acknowledging that wind caused the branch to break, you could still question why it happened just at that particular moment, as you were walking by. This infinite line of causative reasoning is not always so very pronounced. In fact, few people would spend the remainder of the walk contemplating on why the branch broke at that particular moment. Nevertheless, there are indications that our causative reasoning does become very pronounced when we encounter single events, or remarkable ones, in particular if they affect us greatly (or our kin, friends or group).

For humans, single or remarkable events appear to require additional information. Why is that? Humans seem to have a fair understanding of chance and expected frequencies (Boyer 2003); this would be particularly adaptive for a hunter-gatherer hominid who engaged in foraging. However, if events deviate from expected frequencies (i.e. are remarkable) or their expected frequency is unknown (because we represent them as single events, e.g. death of a relative), our mechanisms for reasoning about chance provide little support. While these mechanisms may provide a satisfactory ‘these things happen’ in many cases, they clearly cannot help in interpretation of single of remarkable events. Boyer describes a typical example thereof: “British anthropologist E.E. Evans-Pritchard is famous for his classic accounts of the religious notions and beliefs of the Zande people of Sudan. […] One day, the roof of a mud house collapses in the village where Evans-Pritchard is working. People promptly explain the incident in terms of witchcraft. The people who were under that roof at the time must have powerful enemies. With typical English good sense, Evans-Pritchard points out to his interlocutors that termites had undermined the mud house and there was
nothing particularly mysterious in its collapse. But people are not interested in this aspect of the situation. As they point out to the anthropologist, they know perfectly well termites gnaw through the pillars of the mud houses and that decrepit structures are bound to cave in at a point. What they want to find out is why the roof collapsed at the precise time when so-and-so was sitting underneath it, rather than before or after that” (Boyer 2001).

Boyer (2003) argues that humans represent single and remarkable events in terms of social interaction. Representation could then take the form of ‘people get what they deserve’: the victim owed an invisible agent (causing misfortune) or vice versa (causing good fortune). Or: the (malicious) invisible agent is a cheater, taking things that should not be his. Such thoughts would be the by-product of a hypertrophied social intelligence, i.e. over-activity of our system for social exchange reasoning and cheater detection; this hypertrophy could simply be a reflection of how much human being depended on each other for survival. This role of hypertrophied social intelligence is supported by two facts: (1) people that explain salient misfortune without mentioning supernatural agents still assume agents are involved; this is typically reflected in the human tendency to look for scapegoats (e.g. witchcraft accusation in western cultures several centuries ago). And (2) the connection between the event and the agent thought responsible is typically expressed in social exchange terms: goods given vs. goods received. Many religious notions (of supernatural agents) may be the product of a hypertrophied social intelligence.

3.1.3 Moral intuitions

True morality developed to enable living in ever-larger groups (in turn adaptive because of inter-group competition). The human brain utilizes a number of different mechanisms to support morality, which include means for recognition of group members, expressing and deducing emotional states, feeling empathy, defining and absorbing moral codes and detecting cheaters. It is important that this in essence comes down to a double-layered system for group altruism: group member recognition combined with feeling empathy provides a ‘base layer’ which qualifies as intuitive morality (e.g. one should not murder, hurt or steal from other group members; one should aid weak or hurt group members\(^1\)). Moral codes only elaborate on this intuitive morality by describing more specific scenarios, thus ensuring that any one group member will truly act as altruistic as the other (for large groups, intuitive morality will undoubtedly differ in strength between members, and therefore cannot ensure altruistic homogeneity).

\(^1\) In addition, part of our intuitive morality could be provided by systems developed in sexual selection. For instance, if one’s husband performs adultery, one is less sure he will support you and your children; if one’s wife performs adultery, one invests in costly child support while the child may not be own offspring. Such principles could easily lead to subconscious rules such as: ‘others – i.e. my partner – should not perform adultery’. Ironically, performing adultery oneself would still enhance one’s fitness (for the male: increase gene propagation, for the missus: more sources of child support). Only when such rules made it into the shared set of moral codes, they would apply to oneself as well as to everyone else (because violations of moral codes can be punished).
How do we involve moral intuitions and moral codes in our decisions? Our intuitive psychology handles most of our planning and decision making by contemplating how our actions would affect other individuals, and particularly how these others would react. The easiest, most straightforward way one could involve moral feelings in these decisions may well be by projecting such feelings on another individual. Moral codes already were provided by particular individuals, and could therefore affect our decisions by placing them either with the original providers (e.g. parents), or with some all-compassing ‘group’ individual (i.e. by representing the group as a single individual with beliefs and intentions, its judgments following our moral codes). For moral intuitions, the need (or tendency) to represent these as feelings and judgments of someone else could give rise to a (supernatural) agent that in effect really is a manifestation of one’s own moral intuitions. This does require some minor tweaking of the default concept of human agency: for an agent sharing one’s own moral intuitions, to know right from wrong in any given situation, this agent should not just have the standard, incomplete set of potentially false beliefs. Rather, this agent should have perfect access to information, thus knowing all facts relevant for any (moral) judgment. This is thought to be the principal cause of the omnipresence of all-knowing supernatural agents.

To summarize: our intuitive psychology has become to play such a major role in our process of decision making, that it is easier to represent one’s moral intuitions as beliefs and judgments of an all-knowing – thus supernatural – agent, than to accept them as private, subconscious intuitions.

3.1.4 Death

It is striking how religions everywhere are in part about death. Nearly every religion maintains quite elaborate views on what happens when persons die (and, perhaps even more common, strict guidelines on how to deal with corpses). The may well be due to the fact that death (or more specifically: a corpse) produces one of the most natural, effortless notions of invisible agents (Boyer 2001; Boyer 2003).

Observing corpses of animals, and – to a lesser extent – anonymous humans presents little problem for the mind: inferences from our intuitive biology are triggered, recognize the object is dead (likely something along the lines of: victim of successful predation): it is not conscious, does no longer have an inner force of animation, and will not live again (section 2.2.2). Case closed. It is a straightforward line of reasoning even young children apply. However, this situation changes if the corpse encountered happens to belong to someone you are familiar with, e.g. a member of your kin or your group, or an enemy. In that case, the brain will recognize (typically the face of) the corpse, and put you through to the department ‘intuitive psychology’: you recollect the beliefs and intentions of that individual, and the details of your relationship with him or her. But now we encounter a conflict: your intuitive biology clearly states the individual is dead and no longer relevant, whereas your intuitive psychology cannot just shut-off, and keeps presenting vivid images of the persons previous and future behavior. In essence, you have constructed a do-it-yourself invisible agent. This could also explain why children have great trouble with questions like ‘what happens when some_known_person dies’, while they understand death in predation context perfectly well.
The ‘after-life’ the dead receive from our intuitive psychology is likely to have given rise to one of the most common religious beliefs: the belief in the continued, invisible presence of dead ancestors, still capable of controlling current events. It may also represent the ideal starting point for more abstract religious concepts, such as the Christian notion of the immaterial, immortal soul.

3.2 On supernatural agents

3.2.1 What supernatural agents are like

Supernatural concepts crawl out of the keyholes everywhere: they are produced by an overactive system for agency detection that perceives human agents everywhere, a social intelligence that attributes any single or out-of-the-ordinary event to a (commonly reciprocal) human and an intuitive psychology that prefers to base decisions on agents rather than on one’s own moral intuitions and that fails to recognize individuals cease to exist when they die. Based on these pathways leading to supernatural agents, what can we expect these agents to be like?

First, supernatural agents share many characteristics of normal human agents (in fact, they sometimes were normal human agents until they died). For instance: they engage in social interaction, either following the standard tit-for-tat social exchange rules (thus are benign, or at least fair) or not (thus are malicious cheaters). They are thought to harbor beliefs and intentions (though their beliefs commonly are complete sets of perfect information, causing their expected judgments to follow our moral intuitions). And if given the time to ponder some more on these agents, people may attribute them even more human features, such as appearance, one or more partners and a family life (take for instance the extremely anthropomorphic Greek gods).

There would be nothing supernatural about supernatural agents if they were fully identical to ordinary agents. Typically, they lack one thing: actual physical presence at the ‘scene of the crime’: they cannot be witnessed performing the acts they are thought to be responsible for. This typically leads to the concept of a supernatural agent that either is invisible (spirits, ghosts), or possesses such powers that it can perform actions at a distance (witches, deities). Thus, supernatural agents typically share most features of ordinary agents, but also possess a select – typically small – set of supernatural qualities.

3.2.2 Why supernatural concepts seem real

Why do people maintain beliefs in the existence of agents with counter-intuitive supernatural qualities? First, the actual presence and visibility of agents matters little to our intuitive psychology; as described in section 2.2.3, this inference module is capable of performing offline interaction, i.e. thinking about other agents in their absence. Our brain is therefore well capable of contemplating properties, beliefs, intentions and judgments of supernatural agents. This cannot be all, however: our brain perceives supernatural agents as real, whereas it readily acknowledges characters such as Santa Claus and the bogeyman are imagined. Boyer argues that supernatural concepts are ‘real’ to the brain because they represent a ‘cognitive
optimum’; they trigger many inferences (i.e. are consistent with many things you know), while violating a few – often one – rule of their human agency domain (e.g. it is a human, but an invisible one; it is a human, but one with perfect access to information). This may well be an ideal combination for any concept: it thus becomes both highly plausible, and intriguing.

3.3 Evolution of religion

Many different paths lead to supernatural concepts. They all typically share one characteristic, however: they involve a mental system that operates outside of its designated domain. Our intuitive psychology is triggered in observations that lack human agents (over-detection of agency), or just to make sense of our moral intuitions. Our social exchange reasoning is activated when encountering out-of-the-ordinary events, even if there is no tangible social interaction going on. What evolutionary changes could have enabled these domain transgressions of our mental machinery? First, it is worth noting that our mechanisms for both intuitive psychology and social exchange reasoning can be regarded exclusively human; primates at most possess rudimentary versions thereof. Consequently, these systems must have developed mainly during the past 5 million years of group-based hunting and gathering. As natural selection favored ever-better intuitive psychology and social exchange reasoning, their fast, extensive development may have inadvertently caused them to handle situations outside their original domains, and thus produce notions of supernatural agents. This seems to be the position of Bering (2002), for instance, who even goes one step beyond to argue part of our intuitive psychology developed into an (adaptive) existential theory of mind module.

However, this theory does not do justice to the typical pattern of domain transgressions in the production of supernatural concepts, and it cannot account for the joint emergence of religion and culture, 30,000 to 60,000 years ago. An alternative theory that does do justice to both these facts is provided by Mithen (1999), who argues evolution favored integration of the different, specialized systems in the brain. As these systems became entangled, we would have acquired the capabilities to involve the different mental systems in situations that do not necessarily call for them. This may have brought numerous advantages: application of our intuitive psychology to animals (as we commonly do to pets) could have brought better understanding of animal behavior, and the combination of our intuitive biology and technological (tool-based) thinking could have allowed for application of natural materials in tools and for weapons designed to take advantage of specific weaknesses of animals (and humans). In addition, the integration of mental systems would have produced humans capable of imagination: the ability to see things in other contexts. This would provide the basis for both culture and religion. After millions of years of further divergence and specialization of mental systems, evolution ultimately favored reintegration of the parts, producing a species capable of thinking in many ways about many things, and – coincidentally – a species prone to create religion.
4 Evaluation

4.1 Why religion is not (only) about…

4.1.1 …existential questions

True, several religions provide answers to questions regarding one’s purpose in life, and the consequences of death\(^1\). Also true, such questions seem to pop up time and time again in human minds; witness the ever-growing interest in philosophy. Combined, these facts have led many to believe that existential questions are the basis of – if not the definition of – religion (Kirkpatrick 1999). However, one may well question whether the need to answer existential questions is (and has been) so urgent that it explains 30,000 to 60,000 years of religion.

Where do existential questions stem from? Many assume such questions are produced by a fear of death, inherent in all humans. One can well doubt, however, whether the human brain indeed encodes for such a general fear of death. In a predator-ridden environment where capture of prey and gathering of food are of the essence, a general desire to avoid death would not suffice. Rather, you need (and came to possess) specific mechanisms to deal with specific problems: how do I distinguish predator from prey, food from rock? What does this individual believe, how can I expect it to behave? Buss (1990) likens the human brain to a computer chess program: just like such programs do not contain the instruction ‘make good moves’, the human brain does not just generally fear death. Such abstract instructions are too general to offer any practical value.

More likely, existential questions are just as much a by-product of brain functioning as religion itself. Given a brain with a potential to engage in infinite lines of causative reasoning, and enough spare time on its hands, one is bound to start questioning the nature of one’s own existence. As human brains became less and less occupied with solving primarily problems as hunger and predation, existential questions would have started to pop up more often; in essence, existential questions may represent somewhat of a luxury problem.

Then why does religion focus on existential questions? Well, it does not. Not primarily, at least. Typically, religious thoughts are activated when people deal with concrete situations (this crop, that disease, this new birth, this dead body, etc.) (Boyer 2003). Only a few religions – most notably, Christianity and doctrinal religions of Asia and the Middle East – actually occupy themselves with general questions surrounding life and mortality (Boyer 2003). These are religions with a long history of intense, near-scientific study, i.e. with many years invested in development and expansion of their theories and doctrines. Thus, the emergence of existential questions (and corresponding answers) in such religions could yet again reflect that such questions are mainly the result of a luxury problem.

\(^{1}\) The degree to which these answers provide satisfaction is debatable; for instance, living a life while envisioning a future of eternal hell would not put many minds at rest. Religion can create as much anxiety as it allays (Boyer 2003).
4.1.2 …morality

Several scholars have suggested religion is mainly about a well-defined, fluidly shared set of moral norms and values. By imposing moral codes, religion would maintain coherence in otherwise unattainably large groups, and qualify as adaptation (Alexander 1987; Wilson 2002). Roes valiantly tried to prove this theory, by showing larger societies are indeed characterized by a stronger belief in moralizing gods (Roes and Raymond 2003). But does this fact prove religion is primarily about morality?

As we saw in section 2.3.4, increases in group size were enabled by a well-developed system for definition and absorption of moral codes; as people came to share – and adhere to – moral norms and values, larger coherent societies became possible. But how does one share moral norms and values? To start with, the human brain comes with a distinct set of moral intuitions (produced by systems for empathy, social exchange, sexual selection, etc.). Shared moral codes should not conflict with one’s moral intuitions; conflicts would decrease the chance of moral codes being adequately absorbed, and – ultimately – well distributed. Rather, sharing morals should begin by incorporating one’s defined and explained moral intuitions in a communicable set of moral codes. In other words: moral codes are ultimately shaped by reflections on one’s private moral intuitions. But as described in section 3.1.3, the brain has difficulty with conscious reflections on moral intuitions; our intuitive psychology more than readily creates an all-knowing agent that incorporates our moral intuitions in its beliefs, intentions and judgments. A supernatural agent is born.

So where does this leave us? Larger societies are characterized by more elaborate, better-communicated moral codes. Communication of such codes requires reflecting on one’s own moral intuitions, which leads to definition of all-knowing supernatural agents. Thus: larger societies are characterized by all-knowing supernatural agents said to impose (or at least take a strong interest in) morality. The facts found by Roes are explained, without implying religion serves solely to create morality.

4.1.3 …mystical experience

Mystical, or religious experiences have been suggested to play a major role in the creation of religion. In fact, some have argued religions are based primarily on mystical experiences of few: the mystics’ concepts would lead to the groups’ religious tradition (Boyer 2003). Various researchers have – supposedly successfully – attempted to pinpoint the neurophysiological substrate for religious experience (d’Aquili and Newberg 1998). This in turn has led some to believe religion must be an adaptation: they take an identified neurological base for mystical experiences as proof that religion has an adaptive genetic origin (Ramachandran, Hirstein et al. 1997). Though this may appear a logical line of reasoning, a number of arguments contradict it.

First, only very few practitioners of religion have themselves had any mystical experience (Boyer 2003). Personal mystical experiences are apparently not a requirement for engagement in religious thoughts and practices. Thus, mystical experience could only actually create religion through cultural transmission: concepts acquired by mystics must appear plausible to other individuals, who then can absorb and transmit them. However, as many
anthropologists will argue, for concepts to appear plausible and relevant, they must first comply with many of the things one already knows and believes (as Boyer would argue: the new concepts should trigger many inferences). In other words: for a mystic’s concepts to be absorbed by non-mystics, the latter must already be partial to religious concepts; at the very least, they should have a ‘religious framework’ already in place. This strongly argues against a major role for mystical experiences: such experiences alone could not create religion, as non-mystics need religious notions upfront to accept culturally offered religious concepts.

In addition, many anthropologists argue that the phenomenon of altered states (e.g. mystical experiences) is intrinsically indeterminate (Boyer 2003): one must possess religious concepts to interpret such experiences as mystical in the first place. This makes sense, as the phenomenology of mystical experience is far from unique; it is in fact remarkably similar to that of anxiety attacks (namely, these share a sense of being overwhelmed or engulfed by the feeling, ineffability, and an actual or impending dissolution of the self-as-object) (Kirkpatrick 1999). Thus, how you interpret such experiences would depend on your current state of mind (including your religious concepts, if any); they will rarely cause one to acquire novel, non-obvious (religious) concepts.

Even if mystical experiences were a defining part of religion, the identification of their neurophysiological substrate does not provide sufficient grounds for the assumption that these experiences (and thus, religion) are an adaptation. First, there need not be a unitary psychological mechanism (with identifiable genetic base) for mystical experiences: they may be produced by a combination of elementary systems that are also involved in more mundane emotional experiences (e.g. awe, wonder, aesthetic experiences and creativity) (Kirkpatrick 1999). And, as I have pursued throughout this study, even an identifiable genetic base does not guarantee a trait is adaptive: mystical experiences may result from unintended use of otherwise adaptive structures. These experiences, just like anxiety attacks, may well be an evolutionary by-product.

4.1.4 …rituals

Religion is commonly associated with rituals. As Boyer (2001) argues: “it is rare that people have concepts of gods or ancestors without prescriptions for particular sequences of actions to be performed at specific times and with the expectation of specific results”. However, rituals are not specific to religion; humans perform rituals at many non-religious occasions. For instance, ceremonies of marriage and promotion are little more than big rituals: elaborate scripts of action and role-playing are involved, apparently without contributing substantially to the outcome (why not just settle for a mass mailing stating ‘X and Y are a married couple from now on’?). Rituals and religion are not intrinsically linked, which makes a treatment of the causes and nature of rituals somewhat out of place in this study. Nevertheless, an understanding of rituals may benefit a more complete overview of religion. Therefore, I will treat some commonly mentioned aspects of religion in the following section.

Rituals are typically invoked on two types of occasions: (1) when performing social exchange with supernatural agents (e.g. sacrifices in exchange for protection, or good fortune, crops or health), and (2) when one or more individuals go through a change that has impact on
the group as a whole (e.g. promotion, marriage, births). But what makes these two types of occasions so special that they require special ceremonies?

It is not difficult to understand why social exchange with supernatural agents requires some special actions. Supernatural agents are thought to engage in social interaction (because people attribute single and remarkable events to them), yet are not easily accessible; they are either invisible of acting remotely. Since supernatural agents engage in social interaction, one could charter their aid or protection by providing them with ‘favors’ (e.g. sacrifices). However, the means to achieve this will always be somewhat unconventional as the reciprocal party is either absent or invisible.

Boyer argues that invocation of ritual for group-affecting events also is quite understandable. Changes in individuals can have a major impact on the group. For instance, as a love relationship grows more intimate, other group members should acknowledge the involved individuals are no longer on the ‘mate market’; as a child is born, the priorities of the parents may shift from group interaction to parental care; as boys grow up to be men, their position in the group changes because they obtain other rights, duties, and responsibilities. Yet, such important changes will not automatically be apparent to the whole of the group. They can be too gradual or too private to be noticed. This would be the reason for invocation of rituals such as baptism, marriage and initiation: the change is openly announced to the group, thus allowing others to account for the change, and modify their behavior appropriately (Boyer 2001).

Above facts explain the reasons for distinctly marked ceremonies, but they explain little of the nature of rituals: why do rituals always involve stereotype scripts of action, particular (‘sacred’) objects, role-playing, and feelings of urgency (Boyer 2001)? Kirkpatrick (1999) mentions that ritualistic actions often reflect psychological mechanisms for reasoning in particular domains. For example, magic rituals typically involve actions that are imitative (‘like produces like’), repetitive (‘events which have been observed to occur simultaneously or to follow a particular sequence will continue to follow the same pattern’) or contagious (‘things which have once been together must forever afterward … have a magical effect on one another’) in nature. Each of these seems to reflect evolved inferential rules of intuitive physics and biology (Kirkpatrick 1999). Boyer argues that themes in magic rituals are often very similar to those of Obsessive-Compulsive Disorder (OCD). These themes include: (1) dangerous elements or substances are invisible, (2) any contact (touching, kissing, ingesting) with such substances is dangerous, (3) the amount of substance is irrelevant (e.g. a drop of a sick person’s saliva is just as dangerous as a cupful of the stuff). Such inferences are normally produced by systems for contagion prevention (for viruses and bacteria, all three rules apply), which activate when perceiving danger, or noxious compounds. Thus, ritual themes (and OCD) may in part stem from (over)activation of danger-detection or contagion systems. Many practical aspects of rituals could then be truly spectacular cultural by-products of neural function.
4.2 So where does this leave us?

4.2.1 To the critics

By-product theories of religion provide a viable alternative to adaptationist approaches. Agreed, evolutionary psychology accounts of both evolution of the brain and, to a greater extent, the emergence of religious thoughts and practices can be deemed somewhat speculative on occasions. To their merit, however, by-product theories are capable of painting a rather complete picture of the general aspects of religion. They shed light on many features associated with religions, from supernatural agents to morality, existential questions to sacrifices, obsession with death to rituals. Compared to adaptationist approaches of religion, this is the true strength of such theories: one single all-comprising theoretic framework with some weaker (or less-studied) patches is at least as valuable and plausible as tunnel-vision theories explaining just one aspect with solid evidence.

4.2.2 Benign or evil?

By-products by definition are not adaptations; the former were preserved in evolution as incidental consequences of other adaptive structures. However, this does not imply that by-product traits are neutral to one’s fitness: they can significantly affect fitness, in a positive as well as in a negative sense. Authors of by-product theories often seem partial to the latter: they distinctly qualify religion as maladaptive. For instance, Pinker (1997) first terms religion, together with arts and music, as “Sunday afternoon projects of little adaptive value”, then proceeds to state “Religions … often serve the interest of the people who promulgate them … Priests are wizards of Oz who use special effects, from sleight of hand and ventriloquism to sumptuous temples and cathedrals, to convince others they are privy to forces of power and wonder”. Pinker appears to denounce religion as maladaptive: it is both a waste of time, and a cause of undue subservience. While his statements are extreme in the field of EP, his views seem characteristic of others. For example, both Boyer and Kirkpatrick appear somewhat partial to Pinker’s views, though they refrain from similar strong ridicule.

Qualifications of religion as maladaptive are far from common ground in the whole of psychology, however. On the contrary, many studies have shown considerable health benefits of religion in coping with life’s events. Religion has been shown to produce myriads of beneficial effects: it reduces the time needed to recover from open heart surgery, increases one’s experienced quality of life, and reduces depression and anxiety (Pargament 2002). Apparently, religion can fairly directly increase well-being, and therefore might well increase fitness.

How could religion increase well-being? George (2002) names four means through which religion could benefit health. First, religious participation would promote good health habits, which in turn could have positive effects on health and longevity. This makes sense: religion offers health habits typically in the form of rituals, which – as we saw in section 4.1.4 – likely can be characterized by (over)activation of our systems for contagion prevention.

---

1 That is: maladaptive if taken as a whole; a select few benefit at the expense of the general public.
Second, religion, especially attendance at religious services, would increase social support; this could in particular benefit those that experience little social support to begin with, e.g. minorities and the elderly. That religion could have the capacity to increase social support seems not surprising: religious groups could simply constitute a strongly altruistic in-group, obtaining coherence from a well-defined membership criterion (affiliation, and more prominently, attendance at services) and elaborate, easily transmitted moral codes. Third, George mentions religion could enhance psychosocial resources (e.g. self-esteem, self-efficacy and mastery), which appear strongly related to socio-economic status and social support (and thus, to the religious intimate in-group). And fourth and finally, religion could increase the sense of coherence or meaning: a worldview with three components: meaning, predictability and manageability. This seems to stroke with the notions of supernatural agents as produced by our social exchange systems: events are placed in a context of social exchange with the supernatural, which makes them meaningful and predictable (e.g. they are retributions) and manageable (the agents follow tit-for-tat rules, thus can be influenced). Clearly, the facts do not warrant the idea religion is purely maladaptive; it can bring a bounty of benefits.

Religion is no guarantee for good health, however. The actual beneficial impact of religion is thought to vary by the person, the context and situation, the degree to which religion is well-integrated, and the kind of religion (Pargament 2002). And perhaps most prominently, the efficacy of religion depends on the goals that motivate religious participants: those motivated by goals intrinsic to religious life (which in my opinion can be taken as: those whose religion truly matches their feelings and state of mind) benefit, whereas those motivated by values extrinsic to the character of religion (e.g. status, self-justification) experience negative health effects. Alas for non-believers: religion is no last resort for good health.

4.2.3 Religion and science: distant cousins?

Interestingly, Guthrie (1993) compares religion to science, and concludes they are on an abstract level very similar: they are attempts to interpret and influence the world in general. Both science and religion draw on a framework of observation, logic, analogy, metaphor, model, and unspoken assumption in an attempt to make the world coherent. In fact, it has been argued that science might be regarded as a natural result of religion: “To sum up: the rise of science was not an extension of classical learning. It was the natural outgrowth of Christian doctrine: Nature exists because it was created by God. To love and honor God, one must fully appreciate the wonders of his handiwork. Moreover, because God is perfect, his handiwork functions in accord with immutable principles. By the full use of our God-given powers of reason and observation, we ought to be able to discover these principles” (Stark 2003). Christianity thus could have provided the onset of science.

No clear criteria of evidence, logic, or certainty separate religion from its supposed antithesis, science. Instead, they are separated most strongly by their attitude towards anthropomorphism: science tries to avoid it, while religion takes it as foundation (Guthrie 1993). Whereas religions are inclined to attribute everything to the most complex, detailed
object known to man – a human agent – science aims to exclude all human-like influences, instead focusing on the basic, simple mechanisms\(^1\). Nevertheless, the gap between religion and science may not be as large as often argued: they in essence represent two ends in the continuous spectrum of man’s search for reason and logic.

\(^1\) Thus, science might be regarded an extension of our intuitive physics, religion as an extension of our intuitive psychology.
5 References


