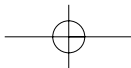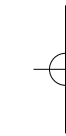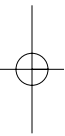# Part V

COGNITIVE ARCHITECTURE
OF THE IMAGINATION

11

Specialized Inference Engines as Precursors of Creative Imagination?

PASCAL BOYER

Abstract.   We usually consider imagination in terms of its high-end, creative products such as literature, religion and the arts. To understand the evolution of imaginative capacities in humans, it makes more sense to focus on humble imaginations that are generally automatic and largely unconscious, and help us produce representations of, for example, what people will say next, that people exist when out of sight, or what aspects of our environment are potentially dangerous. These examples suggest that there may not be one faculty of imagination but many specialized 'what if' inferential systems in human minds.

INTRODUCTION

HOW DOES IMAGINATION WORK and how did it evolve? We have intuitive and seemingly plausible answers to these two questions. We tend to assume that imagination works mostly by suspending some principles that normally constrain inferential processes. We also assume that imagination evolved because it allowed organisms to transcend the limits of the here and now and to create entirely new objects and techniques.

Unfortunately, these two assumptions are not altogether compatible, at least not in this vague form. First, organisms evolve by developing small increments that provide efficiency gains in solving particular problems, not by hitting upon a grand new strategy for dealing with *any* odd problem (Williams 1966). Second, suspending inferential principles should result not in novel insight but in the proliferation of irrelevant associations, hardly an efficient strategy (Tooby *et al.* 2000).

Part of the problem may be that we usually consider the hallmark of imagination as *creativity* or *originality* — and we then try to work downwards, as it were, from such grandiose and exceptional achievements to more

common underlying cognitive processes (Boden 1991, 1994). But human imagination does not just produce high flights of creative fancy. It is also involved in the production of highly stable and fairly predictable representations of possible situations. In some domains of experience, human minds seem automatically to suggest 'what-if' alternatives to current or past experience. For instance, faced with a negative outcome, we cannot help imagine what would have happened had the actual conditions been slightly different (Roese and Olson 1995). We cannot look at a tool without figuring out in what way and for what purpose it may be used (Grafton *et al.* 1997). We interpret certain aspects of the environment as signs of other agents' presence, natural or supernatural (Barrett 2000). We are often led to think of potential dangers in our environments for which there are only indirect cues (Fiddick 2004). In such situations, the production of 'what-if' scenarios is largely outside conscious control and deliberation, although the results, in the form of possible outcomes, are consciously experienced. So, when we consider the evolution of imagination, it may make sense to start with this humble capacity, at the low end of the imagination spectrum, and then see to what extent it is involved in the creativity that we find characteristic of human imagination (Ward 1995).

Considering this partly tacit form of imagination may be of help in addressing two central questions about human imagination, concerning *domain-generality* and *premise-generation* respectively. The question of domain-generality is this: Can we describe imagination of this kind as the output of a particular *faculty*, a functionally distinct system that handles counterfactuals and their consequences? Or is it, alternatively, the output of several specialized 'what-if' inference engines, each of which is dedicated to a particular domain of occurrences and possible scenarios? These two accounts of productive imagination rely on different assumptions about the neuro-cognitive processes involved and carry different implications for what could be expected from productive imagination.

The question of premise-generation is this: are there any constraints on the processes that create the first steps (the premises) of imaginary activities? A very attractive account of imagination depicts it as a random generation of thoughts, followed by the selective retention of those thoughts that support enough inferences to constitute 'viable' imaginary scenarios. Or is there an early selection of imagination premises? Again, these alternative accounts imply very different evolutionary and functional processes.

PREMISE GENERATION: A LATE SELECTION MODEL

We have a great number of studies about the way a mind handles imaginative productions, once it entertains some counterfactual assumption (e.g. *I am a kangaroo*). But we know little about the processes that make some such assumptions (*I am a giraffe*) much more likely than others (*prime numbers are mischievous*).

To illustrate how this issue of premise generation is both central and generally ignored in cognitive models of imagination, consider the closely related domain of pretence. Pretence starts in young children some time around 18 months. It involves typical behaviours such as behaving towards an object (a cardboard box) as if it was another one (a car), and in later developments the child behaving as though they were an animal or another person. There are at least three different computational accounts of pretence, in terms of simulation (Gordon and Olson 1998), propositional attitudes (Leslie 1987) or hypothetical reasoning (Nichols and Stich 1999). These models are accounts of *pretence management*, as it were, rather than *pretence creation*. Our cognitive models of pretence combined with relevant neuroimaging and neuropsychological evidence reveals some fundamental ways of processing counterfactual information.

This leaves aside the question, how these pretence premises are generated. This is a difficult question, notably because it is difficult to formulate in a precise and tractable way. Perhaps the best way is to start with obvious observations about the occurrence of pretence in children. Pretence is an early development and is highly principled, as I stressed above, but it is interesting that it is also somewhat constrained in its themes and targets. Children generally start by pretending that one common, accessible object (a plastic box) is a more interesting but inaccessible one (like a truck), or that a person-like object (doll, toy animal) really is animate; they also develop animal pretence, behaving like their idea of a tiger or dog or elephant; later they also pretend with social roles, acting like doctors, shopkeepers, etc. We know very little about these limits, because our cognitive models operate downstream, as it were, from premise generation.

There is one very tempting way of solving this question—which I would call 'psychological selectionism'. The assumption would be that there are *no* principled processes that generate imaginative premises. That is, the generation of such initial premises is a random process of thought associations produced by the simple relaxation of constraints ordinarily imposed on trains of thought and inferences. Under this account, it would seem difficult to explain why human imagination tends to go back to the same themes, over and over again, and eschew some possible conceptual combinations. But the difficulty could be solved by the other half of a Darwinian model, the process of

242 *Pascal Boyer*

selective retention. Although there are random mutations of thoughts that produce many possible imagination premises, they are not equally 'fit' in the human mind (Campbell 1960). This is what I would call a *late-selection* model, in the sense that the selection of particular contents occurs, in functional terms, after a generation phase during which there is no constraint on what thoughts are entertained.

This is a very tempting hypothesis, not just because it is clearly parsimonious, but also because at least some phenomenology of imagination seems compatible with it. Hypgnagogic states for instance often include associations that are extraordinarily difficult to express once the mind is fully conscious again. One entertains associations such as *my sister really is a prime number*, *electricity is parallel to cheese* which lead to other thoughts of the same kind—unless, that is, one becomes sufficiently awake to consider them, at which point their very content seems to evaporate. The same occurs in various altered states of consciousness induced by either 'disciplines' (e.g. meditation, self-hypnosis) or the use of drugs. Importantly, even in such states, not all associations are that elusive. From dreams or hallucinations or semiconscious states we can also retrieve associations such as *my sister just turned into a huge lobster* or *it is raining needles* that are perfectly fine imagination premises. So we do sometimes experience both quasi-random generation of thoughts of different 'fitness' and a subsequent selection of those that are 'fit' for further processing.

RELEVANCE AND EARLY SELECTION

However, the mutation-selection account may be insufficient. First, imagination-premises are rarely as diverse as a random generation process would imply. Indeed, hypnagogic states and altered states of consciousness are remarkable precisely because they include thoughts that do not stand the test of processing, as it were. In most other contexts, the imagination-premises that become accessible to conscious inspection are all fit for processing. Second, it is difficult to explain how random generation and selective retention would result in the recurrent themes that we observe in such domains as religious imagination or literary fiction, and in fact in most domains of human imagination, as we will see below.

These limits on the range of recurrent human imagination would suggest that the selection of imagination premises operates, not in terms of formal processing capacity, but in terms of the connections between those premises and background knowledge. One may speculate that imagination premises survive cycles of processing and transmission to the extent that they generate novel inferences, relative to semantic memory; and that these novel inferences

would be all the more likely to be retained as they require less processing effort. Understanding imagination-premises in this way, in terms of relevance (Sperber and Wilson 1995), would certainly begin to explain some of the differences mentioned above. Some possible imagination premises (e.g. the artefact without physical existence) support fewer inferences than other ones (e.g. the person without physical existence). For instance, an invisible person may have designs on you, may be watching you, may form moral judgements about you, may be an ally or enemy, etc.; an invisible object is hard to detect . . . and that's that. This difference may explain the different cultural distributions of these two types of concepts (Boyer 2000). So it would seem that a relevance-based model could in principle account for some recurrent themes of human imagination, in terms of a selective process that links them with non-imaginary assumptions.

But the relevance interpretation implies that the generation of imagination-premises (and not just their selection) is crucially dependent on prior representations in semantic memory. Relevance is not just a function of novel inferences generated but also of the processing effort required to activate assumptions that would support these inferences. This is where the organization of semantic memory has a direct effect. To remain in the domain of religious imagination, why is it that non-physical spirits are universal while punctuate spirits (e.g. that only exist on Wednesdays) are never considered? A simple answer would be that it takes great effort to represent the concept of a person (from which a spirit concept is derived) and remove the feature of temporal continuity from its representation. In other words, a feature of semantic memory (that persons are continuous, not punctuate) makes the *generation* of imagination premises with punctuate persons very unlikely.

In the following I will follow this strategy in the description of several domains of imagination, arguing that there may be an *early selection* process that makes certain conceptual combinations much more likely than others. This pre-selection process stems from the fact that in most domains of knowledge we routinely process 'what-if' scenarios that produce counterfactual assumptions and inferences from these assumptions. These spontaneous, generally non-controlled imaginings, in my view, constitute templates on which more deliberate and creative imagination then builds more original scenarios that undergo creative selection. If this is the case, if most knowledge domains include some what-if inference engine, it would follow that these inference engines are independent and specialized—in other words that there is not one imagination but many specialized imaginations.

244 *Pascal Boyer*

GODS, DEAD PEOPLE AND IMAGINARY FRIENDS

Representations of supernatural agents are universal in human cultures, with very similar features (Boyer 1994a) and may have been present from the earliest stages of the cultural explosion in human evolution (Burkert 1996; Mithen 1996). In the past 15 years, various accounts of specific features of religion have converged to constitute what could be called a common or 'standard' model of religious thought and behaviour, based on the notion that religious concepts are a by-product of ordinary cognition (Lawson and McCauley 1990; Boyer 1994b, 2001; Barrett 1996; Pyysiainen 2001; Atran 2002). They 'parasitize' cognitive structures that evolved for other, non-religious reasons. This presupposes a 'selectionist' view of human cultural evolution, as presented in theories of meme-transmission and cultural epidemics (Boyd and Richerson 1985; Sperber 1996). Religious concepts and norms that we find widespread in human cultures are those that resist the eroding, distorting influence of individual transmission better than others.

The standard account stipulates that there is a limited catalogue of supernatural concepts, a subset of which is found in religious systems. Supernatural concepts (found in religion but also in fantasy, dreams, 'superstitions', etc.) are informed by very general assumptions from domain concepts such as PERSON, LIVING THING, MAN-MADE OBJECT. A spirit is a special kind of person, a magic wand a special kind of artefact, a talking tree a special kind of plant. Such notions are salient and inferentially productive because they combine (i) specific features that violate some default expectations for the domain with (ii) expectations held by default as true of the entire domain (Boyer 1994a; Barrett 1996). These combinations of explicit violation and tacit inferences are culturally widespread and may constitute a *memory optimum* (Barrett and Nyhof 2001; Boyer and Ramble 2001). That is, they are better recalled than conceptual combinations that are either more outlandish or more in conformity with expectations.

A subset of this supernatural repertoire consists in religious concepts proper, which are taken by many people as, firstly, quite plausibly real and secondly, of great social and personal importance. These concepts generally describe *intentional agents* so that all standard agency assumptions are projected onto them (Lawson and McCauley 1990). These agents are generally construed as combining (i) explicitly transmitted, counter-intuitive physical or biological features (going through walls, being eternal, born of a god's thigh, etc.), together with (ii) tacitly projected, standard mental features (having perception, memory, intentions and beliefs). In other words, concepts of gods and ancestors require minor but consequential 'tweaking' of standard theory of mind (Barrett and Keil 1996). In this sense, the anthropomorphism that is widespread in religious concepts (Guthrie 1993) is also extremely selec-

tive. The domain of human features that is invariably projected is *intentional agency*, more frequently and more consistently than any other type of human characteristics.

A similar and related projection is found in the familiar but under-studied phenomenon of interaction with dead people. As Jesse Bering and colleagues have shown, common intuitions about the dead, especially the recently dead, are independent from explicit statements of beliefs in the afterlife (Bering 2002). That is, people's judgements about whether, for example, a dead person can think, remember, etc., are only partly predicted by religious or meta-physical commitments. A strong intuition seems to be that mental states seem all the more likely to be projected onto the dead, the more 'purely mental' they are and the more distant from physiology. So the dead are described as *thinking* and *remembering* more often than *feeling*, which occurs more often than *being hungry* (Bering 2002). Indeed, a common feature of interaction with the recently dead is a discrepancy between biological intuitions, which have no problem representing the cessation of function associated with death (Barrett 2005), and our social intuitions which do not stop with biological death (Boyer 2001): thus the recurrent phenomenon of being angry with the dead or proud of them as if they were real social partners.

Another salient case of a common domain of productive imagination is the frequent creation of imaginary friends by young children. From an early age (between 3 and 10) many children (perhaps more than half of them) engage in durable and complex relationships with such agents (Taylor 1999 and Chapter 4). These imagined persons or personified animals, sometimes but not always derived from stories or cartoons or other cultural folklore, follow the child around, play with her, converse with her, etc. Taylor's studies show that having long-term relationships with non-existent characters is not a sign of confusion between fantasy and reality (Taylor and Hort 1990). Young children know perfectly well that their invisible companions are not 'there' in the same sense as real friends and other people.

SOCIAL INTELLIGENCE AND THE ABSENT AGENT

Why is it that a significant part of social interaction takes place with either imagined supernatural agents or actual but deceased agents? Perhaps this feature of spontaneous human imagination is less surprising given human capacities for 'mind-reading' or 'theory of mind', geared to interpreting other agents' (or one's own) behaviour in terms of goals, beliefs, memories and inferences (Leslie 1987; Perner 1991; Whiten 1991). Experimental studies of mindreading, especially of its developmental aspects, have generally focused on actual interaction. Participants were required to infer unobservable

mental states or future behaviour from observable phenomena such as objects, facial expressions, or previous utterances.

The capacity is of great evolutionary advantage, given the human dependence on social interaction. Humans live in a 'cognitive niche' (Tooby and DeVore 1987), in that they, more than any other species, depend on information, especially on information provided by other human beings, and on information about other human beings. This dependence means that mental dispositions that help maintain rich and flexible representations of others, of their goals and their mental states, is crucial.

However, representations of agents *in absentia* are a constant feature of human thinking. Many, perhaps most, of our thoughts about other people are entertained when they are not around. Memories of what people did or said, as well as expectations, fears and hopes of what they may do, are a constant theme of trains of thought and ruminations (and also the quintessential subject matter of social gossip). It may be a special feature of the human mind that we can create such representations and more importantly run social inferences about them. It is certainly a central capacity of human thinking, appears early, is universal and distinctive of normal human minds. This capacity to engage in social interaction *in absentia* is probably at the origin of all the other dispositions for simulated interaction described above.

But why should we have this capacity? From an evolutionary standpoint, it would seem that such thoughts take time and energy away from consideration of present people. An explanation may be that thoughts about absent agents are necessary and useful given the computational constraints of social interaction. Any such situation presents us with a whole gamut of possible actions from our partners as well as possible reactions to our own behaviour. Reactions on our part should be fast but also appropriate. The potential cost of mismanagement of social relations is huge for humans, given their dependence on cooperation for survival (Boyd and Richerson 1990; Gintis 2000). Now there is a trade-off between speed and appropriateness, given the complexity of inferences required for even the simplest social interaction (Cosmides and Tooby 1997). What each actor did or said may convey several intentions, to which there may be several possible responses, and so on.

One way to bypass this computational hurdle may be to have a prepared catalogue of possible interaction scenarios. These would be constructed when the other agent is not around, which would allow sluggish explicit inferences and the slow comparison of different scenarios in terms of plausibility. These scenarios would include appropriate responses. They could be tagged in memory in such a way that they can be quickly activated in actual interaction and provide an intuitive guide to apposite behaviour.

There is some preliminary evidence for the preparation and use of such scenarios in actual social behaviour (Malle *et al.* 2001; Saarni 2001). To focus

on one of the domains mentioned above, Taylor notes that the relationship with an imagined companion is a stable one, so the child must compute the companion's reactions, taking into account not just the imagined friend's personality but also past events in their relationship. What the companion does or says is constrained by their personality and must remain consistent and plausible even in this fantastic domain. Also, companions are often used to provide an alternative viewpoint on a situation. They may find odd information unsurprising or frightening situations manageable (Taylor 1999). So imaginary companions may constitute a form of *training* for the social mind, helping build the social capacities required to maintain coherent social interaction (Taylor and Carlson 1997).

These examples illustrate the *early-selection* process I outlined above. That is, imagination premises seem to be constrained in their generation, even before they are entertained and found inferentially rich or poor. Explicit, and apparently quite diverse, religious concepts are rooted in ordinary and recurrent ways of thinking about absent agents, which are informed by largely tacit principles. This 'absent-agent' cognition, in the view proposed here, creates a whole domain of possible scenarios and inferences about non-physically present agents. Some of these scenarios are then taken up by explicit imagination to become religious thought. Note that only *some* of them will do. As I said above, not all concepts of non-present agents are 'good' for such purposes. Some are too distant from ordinary intuition to support inferences (e.g. gods that only exist on Wednesdays), and some are not attention-grabbing. In this view, then, religious and other supernatural thoughts are the outcome of a late-selection of representations pre-selected by everyday, largely tacit processes of productive imagination.

MIMESIS AS AN EVOLUTIONARY ISSUE

To proceed one step further in the possibilities of imaginative creation, why do people spend inordinate amounts of time and energy thinking about agents they know to be pure fiction? Why the obvious emotional investment in these imagined agents?

There are two puzzles about fiction. One is the classical conundrum about reference, that is, the extent to which we activate standard knowledge in order to make sense of fictional situations and processes. Like other forms of counterfactual thinking, fiction requires both a limited suspension and a massive preservation of default assumptions about the world—and different genres can be defined in terms of how these two operations are handled. Terms such as 'suspension of disbelief' do not adequately convey the complexities of this

very special cognitive activity. Most philosophers and critics interested in the issue have traditionally focused on the logical aspects of fiction (Searle 1979).

Another puzzle concerns the motivation for literary fiction, that is, the evolutionary scenarios that could lead to our common need for and enjoyment of fiction. It would seem at first sight that these have little to do with evolution by natural selection. Attention to non-existent situations and people may be a diversion from actual ones; only the latter would seem to matter as far as survival and reproduction are concerned.

Progress in these two questions may require that we consider another striking phenomenon, the fact that literary fiction (either oral or written) tends to focus on a limited number of recurrent themes, despite massive cultural and historical differences. Despite the occasional opacity induced by different norms and practices, the literary productions of other cultures are generally extremely easy to understand and enjoy, in contrast to other domains of culture such as music or food or clothing. The main reason is that it is very easy to identify exotic themes, genres, and rhetorical devices as broadly similar to familiar ones. Why is it that literature is about a small catalogue of themes and situations? It is clear that literary themes are very close to evolutionary concerns, such as protection against predators (or pseudo-predators such as the Big Bad Wolf), parental investment, proper relations with kin and non-kin, the fitness cost of being orphaned and interacting with stepparents (Leoutsakas 2004), and of course cultural identity and mate-selection, as any reader of Henry James and Jane Austen will confirm.

It is only recently that evolutionary anthropologists and cognitive scientists have started to consider this question and thereby to address the issues of logical status and possible adaptive value (Carroll 1995; Storey 1996; Hogan 2003). A useful starting point to survey these models is to observe that for most people in most human cultures, fiction is of great interest by virtue of its *mimetic* properties. (In view of this cultural universal, it is of course rather unfortunate that the fashion in recent literary studies has been a dogmatic denial of the phenomenon. Literary theorists' vituperative refusal to consider that texts could ever refer to anything has led them to ignore the potential contribution of psychological and biological research to understanding the appeal of literary fiction. On the sorry state of current literary theory, see, e.g., Easterlin and Riebling, 1993; Storey, 1996; Turner, 1996; Carroll, 2004.) Also, literary fiction is not just general counterfactual thinking but, overwhelmingly, counterfactual thinking about *persons*. The first feature has been the object of recent cognitive research, aimed at showing that a narrative capacity was central to the development of the modern human mind about 100,000 years ago (Dautenhahn 2003), that it is pervasive in human cognition (Bruner 1991; Turner 1996; Hogan 2003) and that it

appears very early in cognitive development (Ackerman 1988; Fivush *et al.* 1995).One of the universal features of fiction is the construction of *characters*, that is, imagined agents whose personalities are stable enough to provide a source of expectations and inferences along a narrative. This would suggest that creating and enjoying stories is, among many other things, an exercise in speculative intuitive psychology (Zunshine 2003). In a similar way as absent-agent cognition, it may be valuable mostly for the training it provides, for its pedagogical function (Steen and Owens 2001). What fiction provides is an extensive catalogue of persons and situations, in combinations that are varied enough and complex enough to provide putative scenarios for future interactions as well as the means to make sense of past interactions.

FICTION AND INTUITIVE PERSONOLOGY

However, this functional interpretation may be insufficient, at least if we construe 'theory of mind' in the narrow sense used in current developmental research. The term in principle denotes the capacities involved in ordinary 'mind-reading'. But in practice most studies of theory of mind have focused on the fundamental concepts and processes involved in the capacity, e.g. the concepts of 'belief' and 'intention' and 'agent', and their absence or impairment in some pathologies (Leslie 1994; Baron-Cohen 1995).

Fiction certainly requires (and may help train) those basic parts of 'theory of mind', the detection of inferences, the differentiation between actual states of affairs and various persons' representations of those states of affairs. But it is difficult to see this as its main benefit, mainly because such basic capacities are developed very early and in a robust way in normal minds.

Actual social interaction does not just require the fundamental processes of theory of mind, but also a vast knowledge-base about the way persons behave, an intuitive human psychology in the broader sense (Astington 2003). Smooth social interaction generally requires intuitions that amount to adequate understandings of motivation, feeling, memory, emotion, and reasoning in other agents. Consider even the simplest operations of everyday mindreading. They require assumptions about mental functioning, the connections between intentions and actions, the way people estimate their own actions, the way they are motivated by greed or benevolence or spite, the way a disappointing experience can modify their behaviour, and so on. Each of these assumptions is also modulated, in the case of persons we know, by specific parameters that constitute that person's personality.

One might call this domain of knowledge an intuitive *personology*, a tacit set of principled intuitions on how behaviour is generated in human beings

and how these agents differ from one another. This is the domain of intuitive principles that inform people's expectations about others' behaviours. These principles describe both what people are in general and along what dimensions they may differ as individuals. There are few accounts of this domain in psychology. For instance, the intuitive explanation of behaviour used to be described in social psychology in terms of attribution theory (Heider 1958) which is psychologically insufficient (Harvey *et al*. 1981) and is not entirely consistent with more recent findings on theory of mind (Malle *et al*. 2000). Even in personality psychology, what matters are real differences between people, not people's conceptions of these differences, although people routinely construct social psychological and personality-based explanations of others.

A central task in narrative comprehension is the attribution of causes to behaviours, in terms of intentions, stable dispositions and contextual factors. Also, an essential tension in fiction is the interplay between pure-observer and pure-actor viewpoints (Todorov 1977). On these two fronts, it would seem that the enjoyment of fiction provides a rich training-ground by providing us with a rich and varied set of exercises in applied personology.

This is why I would suggest that literary imagination, like its religious counterpart, is in large part the outcome of an early selection of themes and templates, a set of implicit constraints on what premises can feed into imaginative consideration. True, literary fiction, especially with the use of literacy, explores fictional worlds often far removed from our common intuitions about persons. However, the genres most commonly enjoyed by most people in most cultures remain very close to these intuitions, while providing a refined set of central or extreme examples. Fiction about exceptional people in normal situations or normal people in exceptional situations may help fine-tune out knowledge-base about the fewer people we actually encounter in social interaction.

VARIOUS DOMAINS OF RITUALIZED BEHAVIOUR

One salient form of cultural and personal imagination is the creation of complicated rituals. Characteristics of ritualized behaviour include stereotypy, rigidity, repetition, and apparent lack of rational motivation. Behaviour of this kind is found in cultural rituals, religious or not; in many normal children's complicated routines; in the pathology of obsessive-compulsive disorders (OCDs); in normal adults around certain stages of the life-cycle. These various behaviours may be different manifestations of a specialized mental system with particular neural correlates (Boyer and Lienard 2006).

Standard criteria for OCD in the DSM-IV include (a) intrusive thoughts that (b) cause distress and (c) are accompanied by ritualistic behaviours that

(d) disturb normal activity and (e) are recognized as irrational by the patient (American Psychiatric Association 1995). Typical obsessions include contamination and contagion (fear of catching other people's germs, to ingest contaminated substances, passing on diseases to one's children or others), possible harm to others or to oneself (e.g. handling kitchen utensils and wounding people), as well as social ostracism following shameful or aggressive acts (thoughts about assaulting others, shouting obscenities, exhibitionism, etc.). Most patients are aware that their obsessions are unreasonable and their rituals pointless but report that neither is easily controlled. In most cases the obsessive thoughts are accompanied by rituals supposed to allay anxiety. Some patients engage in bouts of washing or cleaning tools or utensils. Others verify that they locked their door, rolled up the car window, or turned off the gas knobs. Still others are engaged in constant counting activities or need to group objects in sets of particular numbers.

Most normal children develop ritual-like behaviours at some point between the ages of 2 and 10. They report concerns about dirt and cleanliness, and demonstrate a preoccupation with just-right ordering of objects, preferred household routines, repeated action over and over or a specific number of times, strongly preferred food, rituals for eating, awareness of minute details of one's home, hoarding, bedtime rituals.

PRECAUTION SYSTEMS AND THEIR MIMICRY

Some clinical psychologists have suggested that OCD is a perturbation or exaggeration of functions to do with protection from fitness threats (Mataix-Cols *et al.* 2005). The major domains of obsessiveness correspond to particular dangers of evolutionary significance (contagion, harm to infants, social offence, hoarding of precious resources). Szechtman and Woody (2004) put forward a general neuropsychological model of OCD pathology centred on what they call the 'safety motivation system'. The system is present in all normal human beings and monitors external clues of danger. Note that the system deals only with clues of *potential,* not currently detectable, danger. Security-motivation is distinct from the various fear-mechanisms that handle *actual*, currently present danger.

This makes sense of the particular neural structures involved in the pathology. Like Tourette syndrome and some forms of attention-deficit disorders, OCD corresponds to a specific dysfunction of the basal ganglia and of their connectivity, especially of cortex–striatum–thalamus–cortex loops. These structures are involved in the construction of motor 'chunks' or action routines so that a failure of inhibitory connections from the striatum is probably involved in such behaviours. Cortical structures are also involved, in

particular the anterior cingulate and orbito-frontal cortex (OFC). Given the anterior cingulate's role in the detection or prevention of errors and the OFC's role in inhibition, it would seem that OCD pathology combines a hyper-vigilance to possible cues for required motor action in the environment with a difficulty in inhibiting reactions to such cues, even when they prove to present a minimal threat.

If this model is valid, the kind of imagination deployed in the creation of individual rituals (by children and patients) or collective ceremonies seems largely constrained by the kinds of actions and action-sequences identified by a prior Precaution system. It is of course perfectly possible to create many other kinds of ceremonies that go beyond the themes of contagion, predation, and social offence. But it is also remarkable that most rituals in most places tend to revolve around these themes and to use action-patterns that are typical of this prior system.

SPECIALIZED INFERENCE ENGINES

In each of the domains surveyed here, high-end creative elaborations (fiction, religious belief, ritual) seem to be rooted in the more modest achievements of an everyday form of productive imagination. Supernatural agency is an extension of ordinary absent agency; religious and other social rituals are off-shoots of a Precaution system; fiction requires and enriches our everyday per-sonology. In each of these domains, too, we find that ordinary productive imagination stems from the operation of 'what-if' systems or rather a 'what-if' component of diverse knowledge systems. Our ordinary agency systems can function in conditions where the usual stimulus that triggers them (the physical presence of an agent) is removed or replaced with a memory or a generated representation of the agent. In the same way, our personology can be applied to non-existent agents and still deliver constrained inferences about fictional situations. Our Precaution systems go beyond immediate danger to potential threats to fitness that are generally possible rather than actual. Our capacities for handling tools monitor properties of objects (such as affordances) that only exist relative to a particular range of possible goals.

The operation of 'what-if' inferences seems constrained by different principles in the different domains we have surveyed. For instance, the Precaution system only responds to particular cues in the environment and only delivers emotional responses (in the form of slightly or radically increased anxiety) as well as typical responses (disgust, avoidance, cleaning, ordering). Our intuitions about absent agents are couched in the mentalistic idiom of theory of mind, connecting behaviour to beliefs and goals and perceptions. Our intuitive personology takes these mentalistic descriptions of behaviour as

an input and produces causal descriptions in terms of such processes as motivation and reasoning.

Obviously, this does not mean that all creative imagination is bound to follow these well-trodden paths. It only means that there is a strong likelihood, all else being equal, that themes preselected by intuitive inference mechanisms will be found at the basis of creative imagination. In terms of a relevance model, this would be because their processing cost is less than that of manufacturing entirely novel imagination premises (Sperber and Wilson 1995).

IS THERE A COUNTERFACTUAL FACULTY?

Our examples also suggest that the cognitive processes involved are *specific* to each domain. The way a precaution system generates descriptions of potential danger is not at all the way absent-agent cognition creates descriptions of mental states. Different circumstances trigger activation of these different forms of spontaneous productive imagination and the operating principles are different too. This stands in contrast to the traditional way of considering imagination and counterfactuals as the outcome of a central reasoning process, supposedly similar in its operating principles whatever the domain considered. The most precise cognitive accounts focus on counterfactuals, and specify general processes that should be briefly compared to the model outlined here.

In Kahneman's model of norms and comparisons (Kahneman and Miller 2002), each event triggers activation of memories for similar events, which together construct a 'norm' relative to which the current situation is evaluated. The different availability of alternatives is what drives the construction of counterfactuals. The fact that you *could* have avoided this fender-bender by braking a bit earlier is highly salient because the occurrences of braking early enough are highly accessible (Kahneman and Miller 2002). However, the theory does not account for the obvious connection between counterfactuals and current goals. For instance, 'upward' conditionals (in which imagined situations are better than the actual one) seem to be a default while 'downward' conditionals seem to require more effort (Roese and Olson 1997). This is the starting point of Roese and colleagues' 'two-stage' model of conditionals, in which goal-obstruction or failure create a need for counterfactual production, and norms and categories support the content of the counterfactual (Roese *et al*. 2005). These different models predict the production of and reactions to counterfactuals *in general*, not in different domains. The same can be said of most developmental research in this

domain, focusing on children's *general* ability to detach themselves from factual assumptions (Kavanaugh and Harris 2000; Robinson and Beck 2000).

In the current state of this research, it is not possible to say whether such domain-general models can be reconciled with (or can supersede) the domain-specific processes presented here. For one thing, domain-general models are not 'general' in the sense that they have been tested in a variety of cognitive domains and found to have the same explanatory power in all of them. They are general only in the sense that their authors *assume* that similar processes would operate across domains. As a consequence potential domain-differences are rarely explored. (There are of course exceptions. Consider for instance (Tetlock 1998) in the domain of politics, showing how the interaction of world-view premises and outcomes shapes available counterfactuals. See also Fiske and Tetlock (1997) on the limits of moral imagination and moral counterfactuals.)

Classical domain-general models in terms of reasoning are probably insufficient. As Kahneman and colleagues point out, some counterfactuals are produced spontaneously while others are more deliberate and effortful (Kahneman and Miller 2002). This may be because part of the counterfactual production is operated by automatic processes, while the more creative part is under cognitive control (Roese *et al.* 2005). This description in terms of process-dissociation is very close to the model presented here, with both specialized inference systems and explicit, controlled creative imagination. However, once one admits that counterfactuals and imagination are elaborations on the basis of an automatic and constant production of relevant alternatives, it would seem that the latter must emanate from the various systems that produce efficient interpretations of current situations. And since these systems seem to be highly domain-specific, it is plausible to consider that creative imagination too is rooted in very diverse specialized imaginations.

References

Ackerman, B. P. 1988: Reason inferences in the story comprehension of children and adults. *Child Development*, 59, 1426–1442.

American Psychiatric Association 1995: *Diagnostic and Statistical Manual of Mental Disorders* (4th edn). Washington, DC: American Psychiatric Publishing, Inc.

Astington, J. W. 2003: Sometimes necessary, never sufficient: False-belief understanding and social competence. In Repacholi, B. and Slaughter, V. (Eds), *Individual Differences in Theory of Mind: Implications for Typical and Atypical Development*. New York: Psychology Press, 13–38.

Atran, S. A. 2002: In Gods We Trust. The Evolutionary Landscape of Religion. Oxford: Oxford University Press.

Baron-Cohen, S. 1995: Mindblindness: An Essay on Autism and Theory of Mind. Cambridge, MA: MIT Press.

Barrett, H. C. 2005: Cognitive development and the understanding of animal behavior. In Ellis, B. J. (ed.), *Origins of the Social Mind: Evolutionary Psychology and Child Development* : New York: The Guilford Press, 438–467.

Barrett, J. L. 1996: Anthropomorphism, intentional agents, and conceptualizing god. Unpublished PhD dissertation, Cornell University.

Barrett, J. L. 2000: Exploring the natural foundations of religion. *Trends in Cognitive Sciences,* 4(1), 29–34.

Barrett, J. L. and Keil, F. C. 1996: Conceptualizing a non-natural entity: Anthropomorphism in god concepts. *Cognitive Psychology,* 31, 219–247.

Barrett, J. L. and Nyhof, M. 2001: Spreading non-natural concepts: The role of intuitive conceptual structures in memory and transmission of cultural materials. *Journal of Cognition and Culture*, 1(1), 69–100.

Bering, J. M. 2002: Intuitive conceptions of dead agents' minds: The natural foundations of afterlife beliefs as phenomenological boundary. *Journal of Cognition and Culture,* 2(4), 263–308.

Boden, M. A. 1991: *The Creative Mind: Myths and Mechanisms*. New York: Basic Books.

Boden, M. A. 1994: *Dimensions of Creativity*. Cambridge, MA: MIT Press.

Boyd, R. and Richerson, P. J. 1985: *Culture and the Evolutionary Process*. Chicago: University of Chicago Press.

Boyd, R. and Richerson, P. J. 1990: Culture and cooperation. In Mansbridge, J. J. *et al.* (eds), *Beyond Self-interest.* (pp. 111–132). Chicago, IL: University of Chicago Press.

Boyer, P. 1994a: Cognitive constraints on cultural representations: Natural ontologies and religious ideas. In Hirschfeld, L. A. and Gelman, S. (eds), *Mapping the Mind: Domain-specificity in Culture and Cognition*. New York: Cambridge University Press, 391–411.

Boyer, P. 1994b: The Naturalness of Religious Ideas: A Cognitive Theory of Religion. Berkeley, CA: University of California Press.

Boyer, P. 2000: Functional origins of religious concepts: Conceptual and strategic selection in evolved minds [Malinowski Lecture 1999]. *Journal of the Royal Anthropological Institute,* 6, 195–214.

Boyer, P. 2001: Religion Explained. Evolutionary Origins of Religious Thought. New York: Basic Books.

Boyer, P and Lienard, P 2006: Why ritualized behaviour in humans? Precaution systems and action-parsing in developmental, pathological and cultural rituals. *Behavioral and Brain Sciences*, 29, 1–56.

Boyer, P., and Ramble, C. 2001: Cognitive templates for religious concepts: Cross-cultural evidence for recall of counter-intuitive representations. *Cognitive Science,* 25, 535–564.

Bruner, J. 1991: The narrative construction of reality. *Critical Inquiry,* 18(1), 1.

Burkert, W. 1996: *Creation of the Sacred: Tracks of Biology in Early Religions*. Cambridge, MA: Harvard University Press.

Campbell, D. T. 1960: Blind variation and selective retention in creative thought as in other knowledge processes. *Psychological Review,* 67, 380–400.

Carroll, J. 1995: *Evolution and Literary Theory*. Columbia: University of Missouri Press.

256 *Pascal Boyer*

Carroll, J. 2004: Literary Darwinism: Evolution, Human Nature, and LIterature. New York: Routledge.

Cosmides, L. and Tooby, J. 1997: Dissecting the computational architecture of social inference mechanisms. *Ciba Foundation Symposium,* 208, 132–156; discussion 156–161.

Dautenhahn, K. 2003: Stories of lemurs and robots: The social origin of story-telling. In Mateas, M. and Sengers, P. (eds), *Narrative Intelligence*. Amsterdam: Netherlandsing Company, 63–90.

Easterlin, N. and Riebling, B. 1993: *After Poststructuralism: Interdisciplinarity and Literary Theory*. Evanston, IL: Northwestern University Press.

Fiddick, L. 2004: Domains of deontic reasoning: Resolving the discrepancy between the cognitive and moral reasoning literatures. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology,* 57A(3), 447–474.

Fiske, A. P. and Tetlock, P. E. 1997: Taboo trade-offs: Reactions to transactions that transgress the spheres of justice. *Political Psychology,* 18(2), 255–297.

Fivush, R., Haden, C. and Adam, S. 1995: Structure and coherence of preschoolers' personal narratives over time: Implications for childhood amnesia. *Journal of Experimental Child Psychology,* 60(1), 32–56.

Gintis, H. 2000: Strong reciprocity and human sociality. *Journal of Theoretical Biology,* 206(2), 169–179.

Gordon, A. C. L. and Olson, D. R. 1998: The relation between acquisition of a theory of mind and the capacity to hold in mind. *Journal of Experimental Child Psychology,* 68(1), 70–83.

Grafton, S. T., Fadiga, L., Arbib, M. A. and Rizzzolati, G. 1997: Premotor cortex activation during observation and naming of familiar tools. *NeuroImage,* 6, 231–236.

Guthrie, S. E. 1993: *Faces in the Clouds. A New Theory of Religion*. New York: Oxford University Press.

Harvey, J. H., Town, J. P. and Yarkin, K. L. 1981: How fundamental is 'the fundamental attribution error'? *Journal of Personality and Social Psychology,* 40(2), 346–349.

Heider, F. 1958: The Psychology of Interpersonal Relations. New York: Wiley.

Hogan, P. C. 2003: *The Mind and its Stories: Narrative Universals and Human Emotion*. Cambridge; New York: Cambridge University Press.

Kahneman, D. and Miller, D. T. 2002: Norm theory: Comparing reality to its alternatives. In T. Gilovich, D. Griffin *et al.* (Eds), *Heuristics and Biases: The Psychology of Intuitive Judgment* (pp. 348–366): New York: Cambridge University Press, 348–366.

Kavanaugh, R. D. and Harris, P. L. 2000: Pretense and counterfactual thought in young children. In L. Balter and C. S. Tamis-LeMonda (Eds), *Child Psychology: A Handbook of Contemporary Issues*. New York: Psychology Press, 158–176.

Lawson, E. T. and McCauley, R. N. 1990: *Rethinking Religion: Connecting Cognition and Culture*. Cambridge: Cambridge University Press.

Leckman, J. F., Feldman, R., Swain, J. E., Eicher, V., Thompson, N. and Mayes, L. C. 2004: Primary parental preoccupation: Circuits, genes, and the crucial role of the environment. *Journal of Neural Transmitters*, 111(7), 753–771.

Leoutsakas, D. 2004: The orphan tales: Real and imagined stories of parental loss. *Dissertation Abstracts International Section A: Humanities and Social Sciences,* 64(7–A), 2605.

Leslie, A. M. 1987: Pretense and representation: The origins of 'theory of mind'. *Psychological Review,* 94, 412–426.

Leslie, A. M. 1994: Pretending and believing: Issues in the theory of TOMM. *Cognition,* 50(1–3), 211–238.

Malle, B. F., Knobe, J., O'Laughlin, M. J., Pearce, G. E. and Nelson, S. E. 2000: Conceptual structure and social functions of behavior explanations: Beyond person-situation attributions. *Journal of Personality and Social Psychology,* 79(3), 309–326. URL http //www.apa.org/journals/ps html.

Malle, B. F., Moses, L. J. and Baldwin, D. A. (Eds) 2001: *Intentions and Intentionality: Foundations of Social Cognition*. Cambridge, MA: MIT Press.

Mataix-Cols, D., do Rosario-Campos, M. C. and Leckman, J. F. 2005: A multidimensional model of obsessive-compulsive disorder. *The American Journal of Psychiatry,* 162(2), 228–238.

Mithen, S. J. 1996: *The Prehistory of the Mind*. London: Thames and Hudson.

Nichols, S. and Stich, S. 1999: A cognitive theory of pretense. *Cognition,* 74, 115–147.

Perner, J. 1991: *Understanding the Representational Mind*. Cambridge, MA: MIT Press.

Pyysiainen, I. 2001: How Religion Works. Towards a New Cognitive Science of Religion. Leiden: Brill.

Robinson, E. J. and Beck, S. 2000: What is difficult about counterfactual reasoning? In P. Mitchell and K. J. Riggs (Eds), *Children's Reasoning and the Mind*. Hove: Taylor and Francis, 101–119.

Roese, N. J. and Olson, J. M. 1997: Counterfactual thinking: The intersection of affect and function. InZanna, M. P. (Ed.), *Advances in Experimental Social Psychology*, 29, 1–59).

Roese, N. J. and Olson, J. M. (eds) 1995: *What Might Have Been: The Social Psychology of Counterfactual Thinking*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Roese, N. J., Sanna, L. J. and Galinsky, A. D. 2005: The mechanics of imagination: Automaticity and control in counterfactual thinking. In Hassin, R. S., Uleman, J. S. and Bargh, J. A. (eds), *The New Unconscious*. New York: Oxford University Press, 138–170.

Saarni, C. 2001: Cognition, context, and goals: Significant components in social-emotional effectiveness. *Social Development,* 10(1), 125–129.

Searle, J. 1979: The logical status of fictional discourse. In French, P. A., Uehling, T. R. and Wettstein, H. K. (eds), *Contemporary Perspectives in the Philosophy of Language*. Minneapolis: University of Minnesota Press.

Sperber, D. 1996: Explaining Culture: A Naturalistic Approach. Oxford: Blackwell.

Sperber, D., and Wilson, D. 1995: *Relevance. Communication and Cognition* (2nd edn). Oxford: Blackwell.

Steen, F. and Owens, S. 2001: Evolution's pedagogy: An adaptationist model of pretense and entertainment. *Journal of Cognition and Culture,* 1(4), 289–321.

Storey, R. F. 1996: Mimesis and the Human Animal: On the Biogenetic Foundations of Literary Representation. Evanston, IL: Northwestern University Press.

Szechtman, H. and Woody, E. 2004: Obsessive-compulsive disorder as a disturbance of security motivation. *Psychological Review,* 111(1), 111–127.

Taylor, M. 1999: *Imaginary Companions and the Children Who Create Them*. New York: Oxford University Press.

Taylor, M. and Carlson, S. M. 1997: The relation between individual differences in fantasy and theory of mind. *Child Development,* 68(3), 436–455.

Taylor, M. and Hort, B. 1990: Can children be trained in making the distinction between appearance and reality? *Cognitive Development,* 5, 89–99.

Tetlock, P. E. 1998: Close-call counterfactuals and belief-system defenses: I was not almost wrong but i was almost right. *Journal of Personality and Social Psychology,* 75(3), 639–652.

Todorov, T. 1977: *The Poetics of Prose*. Ithaca, NY: Cornell University Press.

Tooby, J. and DeVore, I. 1987: The reconstruction of hominid behavioral evolution through strategic modeling. In Kinzey, W. (ed.), *Primate Models of Hominid Behavior*. New York: SUNY Press, 183–237.

Tooby, J., Cosmides, L., Gallistel, C. R., Fernald, R. D., White, S. A., Sherry, D. F. *et al.* 2000: Evolution. In Gazzaniga, M. S. *et al.* (eds), *The New Cognitive Neurosciences* (2nd edn). Cambridge, MA: MIT Press, 1167–1270.

Turner, M. 1996: *The Literary Mind*. New York: Oxford University Press.

Ward, T. B. 1995: What's old about new ideas? In Smith, S. M., Ward, T. B. *et al*. (eds), *The Creative Cognition Approach*. Cambridge, MA: MIT Press, 157–178.

Whiten, A. (Ed.) 1991: Natural Theories of Mind: The Evolution, Development and Simulation of Everyday Mind-reading. Oxford: Blackwell.

Williams, G. C. 1966: Adaptation and Natural Selection. A Critique of Some Current Evolutionary Thought. Princeton, NJ: Princeton University Press.

Zunshine, L. 2003: Theory of mind and experimental representations of fictional consciousness. *Narrative,* 11(3), 270–291.