Jim Davies and
Jeanette Bicknell

Imagination and Belief

The Microtheories Model of Hypothetical Thinking

Abstract: *Beliefs about hypothetical situations need to be 'quarantined' from factual representations, so that our inference processes do not make false conclusions about the real world. Nichols (2004) argued for the existence of a place where these special beliefs are kept: the pretence box. We show that this theory has a number of drawbacks, including its inability to account for simultaneously keeping track of multiple imagined worlds. We offer an explanation that remedies these problems: beliefs of content imagination each belong to some number of microtheories; systems of ideas tagged as being true or false only in certain contexts.*

Keywords: imagination; aesthetics; philosophy of mind; philosophy of art; hypothetical reasoning; counterfactual reasoning; beliefs.

Introduction

'Imagination' is a slippery term. Leslie Stevenson (2003) has recorded more than twelve different conceptions of imagination, with contexts ranging from the philosophy of mind to fantasy to the effort needed to create enduring works of art. In this paper we examine one aspect of imagination: the ability to generate, analyse, and reason about claims that we recognize or suspect to be false. Implicated in counterfactual thinking, planning, and hypothetical thinking, this kind of cognition is

Correspondence:
Jim Davies, Institute of Cognitive Science, Carleton University, 2202B Dunton Tower, 1125 Colonel By Dr., Ottawa, Ontario, Canada K1S 5B6.
Email: jim@jimdavies.org

necessary not only for creative invention, but even for our ability to understand simple fictional stories.

Our focus, belief-like propositional elements of imagination, can be contrasted with what psychologists call 'mental imagery', which is a sensory-like experience generated by the mind. Imagining that all frogs are purple is a hypothetical belief, but visualizing a single purple frog involves the belief-like entity that the purple frog exists, and then a separate visual representation (perhaps composed of lines and colours) the contents of which do not have truth values. It is only the propositional, belief-like elements that we will address in this paper.

In keeping with much current philosophical and empirical work on the imagination, we endorse the 'single code' hypothesis, which holds that actual and imagined beliefs are represented with the same format, or kind of representation. Not only are beliefs and imaginings *represented* similarly, but they are *processed* similarly as well. Whether the mind is reasoning about how to escape an actual fire, how to escape a hypothetical fire, or the best course of action for a fictional character caught in a fictional fire, it will reason in similar ways, be subject to similar constraints, and come to similar kinds of conclusions.

Despite its explanatory power, the single code hypothesis raises some puzzles. First, we form beliefs about non-occurring, merely imagined situations, and we seem to have no problem keeping these beliefs separate from our beliefs about the real world. For example, when a child engages in pretend play she might 'know' that a banana is a phone in the context of the game, but at the same time have 'decoupled' knowledge that the banana is just a banana in the real world (Lillard, 2001; also called 'double-knowledge' in McCune-Licolich, 1981). We can reason that, if a fire starts in the kitchen, we will be able to escape through the back door. These beliefs about a hypothetical situation (called 'recreative imagination' in Currie and Ravenscroft, 2002) are represented in our minds, and we can manipulate them much like we manipulate beliefs about the real world.

However, the mind can't use these hypothetical beliefs exactly as it does real beliefs, because if we did we would make false inferences about the real world. We usually distinguish actual and hypothetical situations effortlessly (many hallucinations and dreams, while we are experiencing them, are notable counter-examples). If the mind could not distinguish the belief about what to do in case of a hypothetical fire from beliefs about actual fires, we might run out of the house simply because we imagined it was on fire. So while these different

types of beliefs are processed by the mind in similar ways, they must be functionally different in some way.

Another puzzle about the single code hypothesis arises from our engagement with fiction. We form beliefs that are true only in the context of a story or hypothetical situation, and we can keep these beliefs separate from our belief in the objective truth of the story or the hypothetical situation itself. For example, 'Anna Karenina has a small son' is true in the context of the novel, as she is reported as having one young male child. But it is false apart from the novelistic context because there is no such person as Anna Karenina. It is interesting to note that these context-sensitive beliefs can be profoundly moving. Indeed, people seek out particular narratives for their mood-management effects (Green, Brock and Kaufman, 2004). Yet the emotions aroused by fiction typically do not prompt the actions that would be inspired by beliefs about the real world. Readers have cried at the death of Anna Karenina, although none have gone looking for her grave. Such discrepancies have given rise to the 'paradox of fiction' (Radford and Weston, 1975; Pascow, 2004; Schneider, 2013).

People have no trouble inferring counterfactual propositions in known fictional situations (as in *Anna Karenina*). Nor do they have trouble doing this in contexts where the truth of the matter is unknown. For example, someone who has viewed the film *Cool Hand Luke* but does not know who won the Oscar for best actor in 1967 might think to themselves, 'If Paul Newman had won an Oscar for *Cool Hand Luke*, it would have been well deserved' (Byrne, 2005).

In all of these scenarios, hypothetical or context-dependent claims must be treated differently, in some ways but not others, from our beliefs about actualities. How does this work? In this paper we examine a recent influential attempt to answer this question, its difficulties, and then offer a more adequate account.

Nichols' Solution: Pretence Box and Quarantine

Nichols and Stich (2000) introduce the idea of a 'Possible World Box' to explain some of the puzzles associated with children's pretend play. This is a separate mental workspace in which 'our cognitive system builds and temporarily stores representations of one or another possible world' (p. 122). The function of this separate mental space is to quarantine factual representations from the beliefs that arise during pretend play. For example, a child who pretends that a banana can function as a telephone needs to keep this pretence belief separate

from his belief that a banana is a fruit. 'This banana is a phone' is in the Possible World Box.

Within their system, Nichols and Stich include two cognitive mechanisms that further explain our mental lives. The 'updater mechanism' (*ibid.*, p. 124) is the means by which we update our beliefs on the basis of new information, whether this information is perceptual or in the form of a proposition. For example, our belief that 'the weather is fine today' may be revised to 'rain is likely' on the basis of a glance at the darkening sky. It is uncontroversial that we update our beliefs swiftly, reasonably accurately, and largely without conscious effort. On Nichols and Stich's account, the updater mechanism works in the same way for pretences as it does for actual beliefs.

The 'script enabler' (*ibid.*, pp. 126–7) fills in the details of a premise that can't be inferred from the pretence premise itself, from the pretender's set of real-world beliefs, or from her knowledge of what has happened earlier in the imagined scene.

In Nichols (2004) the separate holding area for pretence representations is called the 'pretence box'. In the pretence box, factual representations are kept in quarantine from the belief representations. Consistent with the single code hypothesis, many of the same processes that are usually applied to beliefs can be applied to propositions in this holding area. This allows a person to reason about hypothetical and fictional situations. Inferences made with regard to pretence representations can also somehow draw on beliefs about the real world. So my knowledge about sexual mores informs my reading of *Anna Karenina*, even though I understand that the events relayed by the novel are fictional. However, pretence representations are isolated (in a different 'box') from belief representations. They cannot be used in inferences with beliefs about the real world to generate other beliefs about the real world. Harris (2000) briefly describes a model resembling the pretence box that he refers to as a semi-permeable boundary.

In a later paper (2006), Nichols explores the single code hypothesis at greater length. He considers the question of why it is that belief and imagination seem to have quite different psychological and behavioural consequences. Our affective responses to fictions are less profound than our affective responses to actualities. The axe-murderer on the movie screen is frightening, but not as frightening as the axe-

murderer witnessed in person.[1] Imagination or hypothetical thinking does not typically result in immediate action. Finally, imagination is more subject to mental control.[2] While we can usually distract ourselves from imagined scenarios, even if this can sometimes take real effort, we cannot simply choose at will to reject claims we have previously endorsed. Once we are convinced that global warming is a serious problem, for example, we cannot simply elect to disbelieve it without compelling evidence or arguments. On the other hand, imagining that global warming is not a serious problem is trivially easy.

Problems with the 'Pretence Box' Account of Content Imagination and the Idea of Mental 'Quarantine'

To help introduce the problems with this account consider the following story, which we will refer to as *S*:

> *David was doing his nightly prayers, facing the crucifix on the wall, when he noticed motion near the window. A bat had flown in, and was transforming into a human form, with a vicious smile and sharp fangs. 'This can't be!', David thought, 'Vampires don't exist!'*

Suppose Julie read this story. Reading *S*, her mind forms representations concerning it, which we will refer to as 'imagined beliefs'. David is praying, on his wall there's a crucifix, vampires exist, etc. According to Nichols' and Stich's account, these beliefs are placed in a single special place.

Now imagine that, at the time Julie reads this, she'd been working her way through *The Lord of the Rings* books. She's also built up many beliefs regarding that series, including, for example, the existence of a race of creatures called hobbits. According to the pretence box theory, the belief that vampires exist and the belief that hobbits exist are both in the pretence box, to keep them from contaminating our actual beliefs about the real world.

However, when she goes back to reading the *Rings* books, she will not think that vampires exist in Middle-Earth (the world described in

---

1 There is evidence that when reading stories, though, people spend more time reading it if they think it's fiction than if they think it is non-fiction (Zwaan, 1994).

2 Examples of when imagination is *not* under our control include dreaming, obsessions, anxious thoughts of future problems, or the inability to stop imagining traumatic past events, as displayed in sufferers of post-traumatic stress disorder.

the *Rings* books). At the same time she need not forget that vampires exist in the context of *S*, nor will she believe that hobbits exist in *S*. Somehow she is able to keep the beliefs about Middle-Earth isolated from beliefs about *S*. The problem of quarantine is larger than previous theorists have realized: not only do imagined beliefs need to be kept quarantined from beliefs about the real world, they also need to be quarantined from *other* sets of imagined beliefs! We can call this the 'Many Stories' problem.

We all know that people can keep track of multiple stories, sometimes remembering what is 'true' in those stories for decades. Take, for example, the Shakespeare scholar who can recite the basics of all of Shakespeare's plays at any moment, but never confuses the contents of those plays with each other. But with a single pretence box, the only way a mind can distinguish imagined from real-world beliefs is by whether or not they are in the single pretence box, with no accommodation for an imagined reality *in a particular imagined context*. A single pretence box architecture cannot accommodate our ability to follow more than one story, and Nichols does not offer an account of how we are able to do this.

We need not talk purely of fiction to encounter this problem. We spend a great deal of time imagining ourselves in the past or the future (Buckner and Carroll, 2007), often with different scenarios. Suppose Julie is wondering whether she or her husband should pick up their child at school. She imagines how picking up the child would make her happy, but would interfere with her shopping for dinner. She imagines that if her husband were to pick the child up, he might have to come home early from his book group. If there is only a single pretence box, the two imagined scenarios would be in the same box at the same time, including the following facts: 1) that she will pick up the child, and 2) her husband will pick up the child. This would allow Julie to conclude that in the future both she and her husband will pick up the child. People don't do this. They are able simultaneously to entertain two possible futures without the component beliefs of one affecting those of the other (Johnson-Laird and Byrne, 2002), just as Julie can hold beliefs about the *S* and Middle-Earth at the same time.

One way to salvage the pretence box theory would be to suggest that the box is emptied after Julie reads *The Lord of the Rings* but before she reads *S*. We know this doesn't happen, though, because Julie's Middle-Earth beliefs are not forgotten (as they would be if deleted from the pretence box), nor are they treated as beliefs about the real world (as they would be if they were moved from the pretence box to

the other 'box' of real-world beliefs). Therefore the pretence box account cannot be exactly right. Not only are both stories kept isolated from beliefs about reality, but also they are in isolation from each other. The Many Stories problem is not only unsolved by previous theory, but has not even been introduced as a problem.

A further complication is that beliefs originating from different sources are not *completely* isolated from one another, as a pretence box account suggests. Let's call this the 'Selective Transfer' problem. For example, thinking about *S*, Julie might come to the conclusion that David should grab the crucifix off of the wall to ward off the vampire. Why does she make such an inference? Where does she get the idea that vampires can be warded off with crucifixes? Clearly, she has formed this belief based on yet other (fictional) stories about vampires. She knows about vampire lore, even though she knows that vampires do not exist. She's applying vampire lore, presumably from its own isolated area, to *S*, but not to *The Lord of the Rings*. Also, as we saw above, when reasoning about hypothetical and fictional situations of all sorts, we apply our knowledge of the real world to greater or lesser extents. For example, Julie can infer that David has hands with which he can grab the crucifix, even though hands are not mentioned in *S*. She is drawing this information from her knowledge of people in the real world — knowledge that is not in any pretence box at all.

Nichols and Stich (2000) account for the Selective Transfer problem with scripts, which is a concept borrowed from artificial intelligence. A script is a stereotyped set of ordered actions, such as eating at a restaurant. These are created in the mind through experience with the real world. It's how we know to ask for the bill after the meal is finished. According to Nichols and Stich, we use these scripts to help us make sense of imagined scenarios. However, it is worth noting that scripts are rather limited in what they can do.

Scripts tend to be descriptions of step by step, stereotypical activities in the real world. They are not as complex as full mental models or mental simulations. Nor are they about general semantic knowledge (e.g. 'vampires can be warded off with crucifixes') without a temporal component. They are about 'event-based situations' (Abelson, 1981). Nevertheless with the script enabler Nichols and Stich provide some mechanism for bringing beliefs, of a sort, into the quarantined area. Also, it is not clear that scripts can be formed from other pretences (such as how to kill vampires). Scripts are described as being about the real world, not about fictional worlds, so on the face

of it a script account of the Selective Transfer problem is inadequate. But we might give it a generous interpretation and suppose that scripts can also be made to describe stereotypical activities in fictional worlds as well.

The script enabler theory does not offer an explanation of how a cognitive system would know when to apply scripts to some pretences and not others (*S* and not *The Lord of the Rings*).

Julie's comprehension of *S* is even more complicated. Julie simultaneously knows that vampires exist in the world of *S*, and also that the character David doesn't believe in them. What's interesting about this example is that Julie doesn't believe in vampires either, at least not in the real world. But she knows that, though she might be correct in her world, David is incorrect in his. David's belief, which would be true in the real world, is false in the world of the story.

In summary, we see two fundamental problems with the pretence box theory with respect to imaginings, hypothetical situations, and stories (all of which we will refer to simply as 'fictions'). First, the Many Stories problem: the pretence box theory does not account for people's ability to keep track of different fictions without confusing them (e.g. not confusing *S* and *The Lord of the Rings*). Second, the Selective Transfer problem: it does not account for people's ability to use real-world beliefs to understand fictions, where these real-world beliefs are inferences about human anatomy (e.g. that David has hands), unless they are a part of a script — a part of a larger, stereotyped activity. It also seems unable to account for semantic knowledge in the context of particular literary genres (e.g. using broader vampire lore to infer that David should use a crucifix).

Our solution to these problems begins by discarding the 'box' metaphor. Conceiving of beliefs — whether these beliefs are veridical, hypothetical, or part of an imaginative game — as located in different boxes is unhelpful. We propose a different mapping of mental life, one that we are convinced makes better sense of the logic as manifest in mental states, and in particular of imagination and pretence. Rather than being 'quarantined' in 'boxes', beliefs are organized into microtheories, and these theories are arranged in hierarchies.

The Use of Microtheories in CYC

The microtheory concept is borrowed from the giant artificial intelligence project CYC (short for 'encyclopaedia' and pronounced 'psych'; Lenat and Guha, 1990). The long-term goals of the CYC

project are to explicitly represent all human common-sense knowledge. It is, by far, the largest project of its kind, orders of magnitude larger than any others.

One of the interesting findings of the CYC project was that there are very few propositions that are universally true, that is, true independent of context (*ibid.*). Something that is true in one context is false in another, potentially leading to contradictions in the database. For example, in everyday life water is wet, and when something touches something else that is wet it will tend to get wet itself. Getting wet (in the case of water) means a transfer of water from one thing to another. None of this is true for a water molecule, however. It is meaningless to say that a single water molecule is wet, as wetness is an emergent property of massive numbers of water molecules together. When one water molecule touches another, it does not get wet, nor does water come off of the water molecule and stick to the other one. Nor is it true for water in the form of ice.

Examples abound: mammals are hairy, but marine mammals usually are not. Blue objects appear blue, but not in the darkness, nor under other unusual lighting conditions, and are usually pigmented blue, but not in the case of the sky, or for blue eyes. People tend to believe what they say, but not if they are playing characters in a film, and not if they are lying.

Faced with such examples, the solution of the CYC project was to use microtheories. In CYC, each proposition is tagged with some set of microtheories for which that fact is to be considered true. So, for example, the idea that something touching water will get wet will be true in a microtheory such as ‘what happens in a bathroom’ or on a beach, but will not be true in the chemistry microtheory. A particular proposition might have many microtheories for which it is true. These microtheories can be nested — for example, organic chemistry might be nested within chemistry, which means that everything true of chemistry is true in organic chemistry, but not necessarily vice versa. This is CYC’s solution to its own quarantine problem — propositions from incompatible microtheories may not be used to create new inferences.

Lenat and Guha present the microtheory idea only as a solution to a practical AI problem, and it is not presented as a model of human cognition.

The Microtheories Model of Propositional Imagination

We extend the use of CYC microtheories into the Microtheories Model (MTM) by applying the idea to human cognition, and particularly to hypothetical reasoning.

In the Microtheories Model, each belief, be it about something real or fictional, is a part of some set of contexts, or microtheories. There is no single, generic bin for all hypothetical beliefs. Nor are there a number of smaller boxes or holding places for different scenarios. Imagine the chaos in our mental life if there were a box for each hypothetical situation. Reasoning about what to do if there was a fire in the kitchen would be in a different mental space than reasoning about what to do if a fire started in a public place, and because of their mutual quarantine it is not clear how reasoning about one could inform reasoning about the other. It seems more plausible that our beliefs about what to do in a fire belong to a microtheory, which is modified for different physical locations, times of day, etc. The microtheory will contain veridical beliefs ('Fire is dangerous', which would also be in other microtheories) and hypothetical beliefs ('If a fire in the kitchen blocks the front door, I should head for the basement exit'). Such a microtheory about fire may be connected with related microtheories about other possible household disasters. The part of our mental lives that is devoted to hypothetical reasoning is, in large part, an exercise of thinking through different microtheories, modifying or updating them upon the acquisition of new information, comparing and contrasting them with different microtheories, and joining together elements from different microtheories. All of this is much richer than either a single pretence box or multiple, separate pretence boxes would allow.

Although the microtheories in CYC were not built to accommodate hypothetical thinking, because they quarantine beliefs away from others, they can serve this function in the MTM. We will describe how MTM accounts for human abilities that the pretence box theory cannot: 1) people's ability to keep track of different fictions without confusing them (the Many Stories problem); and 2) people's ability to selectively use real-world beliefs to understand fictions (the Selective Transfer problem).

To return to the examples presented in the introduction, Julie would tag as 'true' her beliefs about the actual world. Other propositions might be tagged as 'true in the world of *S*', or 'true of vampire stories

generally', or 'true of Middle-Earth as described in *The Lord of the Rings*'. In the MTM, a tag represents a belief's inclusion in a particular microtheory. Julie's beliefs about the character David are true, and believed by Julie to be true, only in the context of the story — as mentally represented with a tag indicating that such beliefs are true in the context of the story *S*. Beliefs about vampires and how to deal with them are true in the context of stories about vampires. In terms of the architecture of mind, we would describe each belief as having a truth value only with respect to some set of microtheories. Vampires exist in vampire stories, but not on Middle-Earth, and not on real Earth either. Since Julie has separate microtheories for Middle-Earth and for the vampire story, she will not conclude that vampires exist in Middle-Earth, as she might if those beliefs had been in the same pretence box. To extend the medical metaphor, having only one pretence box is like isolating people with leprosy, ebola, and polio in the same quarantined hospital room. This diversity of microtheories, architecturally implemented as tags, is our response to the Many Stories problem.

Defenders of the pretence box account might argue that because the updater mechanism computes over contents in both the belief box and the pretence box, the Many Stories problem does not arise. They could argue that our belief box contains meta-beliefs about different stories, such that one of my beliefs about *The Lord of the Rings* is that it is set in a different world than that of the story *S.* We agree that something like an updater mechanism must exist. As we engage with fiction and with pretences (just as we engage with the world), our microtheories are revised. And certainly the pretence box presumably holds various meta-beliefs that help us to differentiate the many stories that we track, though this is not explicit in any of their papers. However, to really make sense of our engagement with fiction and of our cognitive abilities more generally, we require not just meta-beliefs about each fiction, but a set of related, revisable beliefs, the truth conditions of some dependent on the truth conditions of others. In short, we require a theory for each story. So for the pretence box model of cognition to make sense, we have to have relevant theories in the belief box, each having scope over some propositions in the pretence box and not others. But the whole point of having a bunch of propositions in the same box is so that they can be treated in the same way, raising the question of why there should be only one box.

Next we consider the Selective Transfer problem — the human ability to use some (but not all) real-world beliefs to make sense of stories. Just as microtheories can be nested in CYC, they can be nested

in MTM as well. The vampire story Julie read has an associated microtheory that is embedded within the microtheory of vampire lore. That is, beliefs about vampire lore are true in *S*, but not every aspect of *S* is necessarily true of vampire lore more generally. A microtheory can be labelled as 'nested' in another, which means that it can inherit beliefs from the nesting microtheory. In general, beliefs in the nesting microtheory can affect beliefs in the nested theory, but not the other way around. So the theories work 'top-down'. This allows Julie to reason that David should grab the crucifix. The vampire lore microtheory is embedded within beliefs about the real world, allowing real-world beliefs to affect beliefs in the microtheory, as in the case where Julie infers that David has hands. (When we create fictional worlds, we assume by default that things true in the real world are true in the fictional one, Walton, 1990; Ryan, 1991; Gerrig, 1993; Thomasson, 1998.) But beliefs in the vampire lore microtheory will not be used to generate new beliefs for the real world. To take another example, if children are engaged in pretence play outside, and it begins to rain, they will often incorporate the change in the weather into their pretend play scenario, rather than ignore it. Children engaged in pretend play will use causal reasoning from the real world to inform the fantasy (Harris, 2000).

Nesting in the MTM accounts for the possibility of selective transfer. Julie's appreciation of *S* is enhanced by her real-world beliefs about the supposed power of a crucifix in combatting vampires. The hierarchical nature of microtheories explains how Julie could infer that David should grab the crucifix. She believes that vampire lore is true in the story, and since all of it is within a microtheory of the real world, she can infer that David has arms, that the Crucifix on the wall indicates he is Christian, etc. In this way beliefs can trickle down from nesting to nested microtheories. But her appreciation of *S* would be diminished if the real-world belief 'vampires don't exist' interfered with her engagement in the narrative. On the pretence box account, she would have to 'jump' between the real-world box and the world of the story in the pretence box, with some (unspecified) way of selecting which real-world beliefs to transfer over to the pretence box.

Nichols and Stich partially address this difficulty with the inclusion of the script enabler mechanism. However, as mentioned above, scripts are poor at representing general semantic knowledge that is not represented as events in a stereotyped sequence. Further, there is no notion of embeddedness or any kind of hierarchical structure either in

the description of the script enabler system (Nichols and Stich, 2000) nor the original script theory that inspired it (Abelson, 1981).

The MTM provides a more elegant solution here. In reading *S* Julie operates under a microtheory that is specific to the story, which is nested within her microtheory of the actual world. While reading, she updates and changes her microtheory of *S.* As her real-world beliefs change, her microtheories of different stories may also change. (We acknowledge that we haven't solved the thorny problem of deciding *which* beliefs are selected for transfer.)

Microtheories Applied to Explain Theory of Mind

Recall that Julie can reason about the mental state of David in *S* when she believed that he did not believe in vampires. In this section we would like to show how microtheories can help explain people's abilities to represent and then reason about others' mental states. If we imagine other people and their beliefs as 'contexts', we can understand and reason about others' beliefs by keeping them in separate microtheories. For example, one might have an orthodox Jewish friend, Sarah, who believes that there will be a Jewish American president in the next 20 years. Others might disagree. Microtheories can keep track of who believes what — something we do every day. Also, one might infer that Sarah also believes that the messiah has not come yet, because she is a member of a Jewish culture that, generally, holds that belief to be true. Sarah's beliefs form a microtheory that can be nested within a microtheory for Jewish beliefs, as well as others (perhaps 'beliefs of people who live in Toronto'). Thus the microtheory idea, as well as their hierarchical organization, can help an agent keep track of people's beliefs.

In our running example, Julie has beliefs about David's beliefs. David's beliefs form a microtheory, nested in the story *S* microtheory, nested in the vampire lore microtheory in Julie's mind. She stores the fact 'vampires exist', but keeps it tagged with the *S* and vampire lore microtheories. She also stores the fact 'vampires don't exist', tagged with the microtheory of David's beliefs, and, of course, the real world microtheory. This allows her to reason without there being contradictions. It also allows her to reason that, in the *S* microtheory, David's belief is false.

In a very complicated story, we might believe that the character Alice believes that Joan believes that Tom believes that the stove is on. This would be accommodated by nested microtheories. Most

people cannot go beyond four orders of mental-state reasoning, but some have been known to go as high as seven (Bering, 2011), suggesting an empirical limit to nested microtheories in human mental architecture.

The implications of MTM go beyond understanding fiction and keeping track of our own plans and imaginings. We discuss CYC's version of the theory to note that it is very probable that beliefs in general, not just those about hypothetical situations, are organized into microtheories. Just as vampires only exist in certain stories, and hobbits in others, when you release an object on Earth it falls, but if the object is released in interstellar space, it might not. Not only do we have multiple, nested microtheories of hypotheticals, but of reality as well.

Currie and Ravenscroft (2002) seem not to accept the single code hypothesis, preferring to refer to fictional beliefs as 'belief-like imaginings'. In another paper (1995) Currie calls them 'make-beliefs'. MTM is more parsimonious, in that it requires only one kind of belief in the mind. The tagging system accounts for imagination, fictional beliefs, plans, and, possibly, theory of mind.

Harris and Kavanaugh (1993) suggest that objects in pretend beliefs are flagged as being imagined. Currie (1995) extends this to suggest that the propositions themselves are what are flagged. Though this superficially sounds like the tagging described in MTM, because there is only one kind of flag, the theory is functionally identical to the pretence box model, and subject to the same problems.

Limitations of MTM and Anticipated Objections

We will anticipate some objections to our approach. The weakest part of our theory is the inherent logic or the system of rules by which beliefs may and may not interact. In our theory, reasoning is top-down. That is, beliefs nested within a specific microtheory may not be used to generate or modify beliefs in its containing supertheory. However, it's clear that this happens. For example, how did Julie get her beliefs about vampire lore in the first place? From vampire stories, of course, as read in books and seen in films. Somehow a prototypical vampire lore microtheory is created from common elements in multiple microtheories. How this occurs is beyond the range of this paper, but we will suggest that the process might resemble how, in general, concepts are abstracted from individual instances. We will call this the 'trickle-up' problem.

Another issue we have not commented on is what Gendler (2008) calls 'aliefs', which are belief-like propositions that our deliberative minds do not endorse. Optical illusions provide a good example. In the Müller-Lyer illusion, two lines appear to be of different lengths, but we can measure them and 'know' that they are not. However, the knowledge gained by measuring does not eliminate the continued perception of the lines being of different lengths. The encoded version of the proposition stating that the lines are different would be an alief, according to Gendler: the early perceptual system 'believes' it but the more reflective, deliberative system does not. Let's take, for example, Charles Bonnet syndrome (Sacks, 2013), which causes visual hallucinations without any auditory component. Some, at first, talk to the hallucinated people. Finding that they are unresponsive makes people acknowledge that what they are experiencing is an hallucination. The MTM explanation of what is happening here is that the initial belief, say 'some people are there', has a 'real-world' tag, but when one does not believe in it anymore that tag gets replaced with an 'hallucination' tag. It seems that aliefs could be used in some inferences and not in others, but neither MTM nor the pretence box account shed light on this issue.

While some elements of isolation exist and are necessary, there is also a good deal of permeability between microtheories beyond trickle-up. Currie and Ravenscroft (2002) suggest that an imagined proposition used with a real-world belief to generate an inference will generate only imagined conclusions. Similarly, our MTM's nesting theory allows beliefs to move from the real world microtheories to their nested fictional ones. But researchers have found that beliefs arising in the context of fiction can alter our beliefs about the real world. We learn a great deal about the real world from fiction, which can be as powerful as factual narratives with respect to changing real-world beliefs (Green and Brock, 2000), including, of course, some false propositions (Marsh, Meade and Roedinger, 2003), and values (Shrum, Burroughs and Rindfleisch, 2005). If our mental lives were really encapsulated in a series of boxes, the contents of which were *completely* quarantined from one another, none of this would be possible. No theory, including MTM, Currie and Ravenscroft's, nor Nichols', can account for the effects described above.

To take another example, we learn about history from James Michener, science from Michael Crichton, and most of what we believe about courtrooms and police come from the films and books we've seen about them. Sometimes we are successful, in that we form

new true beliefs about the real world based on fiction. At other times, however, the beliefs we form on the basis of stories are mistaken. Harris (2000) calls these 'intrusion effects'. How we choose which beliefs to transfer from the microtheory of the world of the story to our beliefs about the real world is a messy problem. Our theory, at this stage, has no specific answer to it. However, it is striking that often people usually have no trouble knowing what beliefs to trickle up to a supertheory and which not to. In *Jurassic Park,* the reader might walk away with new real-world beliefs about DNA, but not new real-world beliefs regarding specific plot events. Perhaps some cultural understanding of the nature of scientific and historical fiction guides the reader in these cases, but as of yet there is no *general* theory for how a particular proposition gets put in a pretence box or placed in a particular microtheory.

A more striking example of hypothetical models bleeding into reality comes from the false memory literature. Loftus and her colleagues found that simply imagining a childhood experience is often enough to report it as having actually happened (Bernstein, Godfrey and Loftus, 2009; Garry *et al.*, 1996). We conjecture that whether or not a belief from a fictional microtheory can trickle up into real-world beliefs has something to do with whether we understand the fiction to be 'based on truth', and the standards we understand that holds to, and whether or not the beliefs in question are consistent with beliefs we already have about the world. However, people often cannot remember if a story they read was fiction or non-fiction (Green and Brock, 2000). It could be that imagining something simply makes it more familiar, which increases believability (Bernstein, Godfrey and Loftus, 2009; Garry and Polaschek, 2000). It also could be that tags connecting beliefs to microtheories, like other mental representations, can simply be forgotten.

It is striking that our beliefs about fictional situations are unlikely to affect our views on reality, but the emotions we feel from fiction are just as real as those we feel from real life (Harris, 2000). Not only does it feel the same (perhaps differing in intensity), but adults reading frightening text have the same physiological reactions as fear from real things, such as changes in skin conductance and heart rate (Lang, 1984). This suggests the interesting idea that propositions can be quarantined but emotions cannot!

We also do not have a brain account of how the MTM might be implemented, though competing theories suffer from this same problem. Little is known about the neuroscience of the relationship

between real-world beliefs and imaginings, but one group of researchers has found a particular brain fold that is implicated in distinguishing imagination and reality (Buda *et al.*, 2011). Dreaming is a major imaginative activity, and critical and planning areas of the brain are subdued during it, perhaps contributing to why we tend to believe in our dreams while we're experiencing them (Mazur, Pace-Schott and Hobson, 1998).

Treatments of these problems are promising areas of future work.

Conclusion

We have presented a theory of belief organization that goes beyond Nichols' and Stich's pretence box theory to better accommodate how veridical and hypothetical beliefs interact in understanding fiction, imagined situations, making sense of others' beliefs, and making sense of reality. Beliefs belong to sets of microtheories, and their nestedness determines how they can interact in an individual's mind.

References

Abelson, R. (1981) Psychological status of the script concept, *American Psychologist*, **36**, pp. 715–729.

Bering, J. (2011) *The Belief Instinct: The Psychology of Souls, Destiny, and the Meaning of Life*, New York: W.W. Norton & Company.

Bernstein, D.M., Godfrey, R.D. & Loftus, E.F. (2009) False memories: The role of plausibility and autobiographical belief, in Markman, K.D., Klein, W.M.P. & Suhr, J.A. (eds.) *Handbook of Imagination and Mental Simulation*, New York: Taylor & Francis Group.

Buckner, R.L. & Carroll, D.C. (2007) Self-projection and the brain, *Trends in Cognitive Sciences*, **11** (2), pp. 49–57.

Buda, M., Fornito, A., Bergström, Z.M. & Simons, J.S. (2011) A specific brain structural basis for individual differences in reality monitoring, *The Journal of Neuroscience*, **31** (40), pp. 14308–14313.

Byrne, R.M.J. (2005) *The Rational Imagination: How People Create Alternatives to Reality*, Cambridge, MA: MIT Press.

Currie, G. (1995) Imagination and simulation: Aesthetics meets cognitive science, in Davies, M. and Stone, T. (eds.) *Mental Simulation: Evaluations and Applications*, Oxford: Blackwell.

Currie, G. & Ravenscroft, I. (2002) *Recreative Minds: Imagination in Philosophy and Psychology*, New York: Oxford University Press.

Garry, M., Manning, C.G., Loftus, E.F. & Sherman, S.J. (1996) Imagination inflation: Imagining a childhood event inflates confidence that it occurred, *Psychonomic Bulletin and Review*, **3**, pp. 208–214.

Garry, M. & Polaschek, D.L.L. (2000) Imagination and memory, *Current Directions in Psychological Science*, **9**, pp. 6–10.

Gendler, T.S. (2008) Alief and belief, *The Journal of Philosophy*, **105** (10), pp. 634–663.

Gerrig, R.J. (1993) *Experiencing Narrative Worlds: On the Psychological Activities of Reading*, New Haven, CT: Yale University Press.
Green, M.C. & Brock, T.C. (2000) The role of transportation in the persuasiveness of public narratives, *Journal of Personality and Social Psychology*, **79**, pp. 701–721.
Green, M.C., Brock, T.C. & Kaufman, G.F. (2004) Understanding media enjoyment: The role of transportation into narrative worlds, *Communication Theory*, **14** (4), pp. 311–327.
Harris, P.L. (2000) *The Work of the Imagination*, Oxford: Blackwell.
Harris, P.L. & Kavanaugh, R.D. (1993) Young children's understanding of pretense, *Monographs of the Society for Research in Child Development*, **58** (1, Serial No. 231).
Johnson-Laird, P.N. & Byrne, R.M.J. (2002) Conditionals: A theory of meaning, pragmatics, and inference, *Psychological Review*, **109** (4), pp. 646–678.
Lang, P.J. (1984) Cognition and emotion: Concept and action, in Izard, C.E. & Zajonc, R.B. (eds.) *Emotions, Cognition, and Behavior*, Cambridge: Cambridge University Press.
Lenat, D.B. & Guha, R.V. (1990) Cyc: A midterm report, *AI Magazine*, **11** (3), pp. 32–59.
Lillard, A. (2001) Pretend play as twin earth: A social-cognitive analysis, *Developmental Review*, **21**, pp. 495–531.
Marsh, E.J., Meade, M.L. & Roedinger, H.L. (2003) Learning facts from fiction, *Journal of Memory and Language*, **49** (4), pp. 519–536.
McCune-Nicolich, L. (1981) Toward symbolic functioning: Structure of early use of early pretend games and potential parallels with language, *Child Development*, **52**, pp. 785–797.
Muzur, A., Pace-Schott, E. & Hobson, J.A. (2002) The prefrontal cortex in sleep, *Trends in Cognitive Sciences*, **6** (11), pp. 475–481.
Nichols, S. (2004) Imagining and believing: The promise of a single code, *Journal of Aesthetics and Art Criticism*, **62** (2), pp. 129–139.
Nichols, S. (2006) Just the imagination: Why imagining doesn't behave like believing, *Mind & Language*, **21** (4), pp. 459–474.
Nichols, S. & Stich, S. (2000) A cognitive theory of pretense, *Cognition*, **74**, pp. 115–147.
Paskow, A. (2004) *The Paradoxes of Art: A Phenomenological Investigation*, Cambridge: Cambridge University Press.
Radford, C. & Weston, M. (1975) How can we be moved by the fate of Anna Karenina?, *Proceedings of the Aristotelian Society, Supplementary Volumes*, **49**, pp. 67–93.
Ryan, M. (1991) *Possible Worlds, Artificial Intelligence, and Narrative Theory*, Bloomington, IN: Indiana University Press.
Sacks, O. (2013) *Hallucinations*, New York: Vintage.
Schneider, S. (2013) The paradox of fiction, *Internet Encyclopedia of Philosophy*, [Online], http://www.iep.utm.edu/fict-par/ [30 April 2013].
Shrum, L.J., Burroughs, J.E. & Rindfleisch, A. (2005) Television's culture of material values, *Journal of Consumer Research*, **32** (3), pp. 473–479.
Stevenson, L. (2003) Twelve conceptions of imagination, *British Journal of Aesthetics*, **43** (3), pp. 238–259.
Thomasson, A.L. (1998) *Fiction and Metaphysics*, New York: Cambridge University Press.

Walton, K.L. (1990) *Mimesis as Make-Believe: On the Foundations of the Representational Arts*, Cambridge, MA: Harvard University Press.
Zwaan, R.A. (1994) Effect of genre expectations on text comprehension, *Journal of Experimental Psychology: Learning, Memory and Cognition*, **20** (4), pp. 920–933.

Paper received May 2014; revised October 2015.