MODEL-BASED REPRESENTATION OF COGNITION

Aaron Bramson

January 25, 2008

Abstract: Propositional representation is the dominant paradigm in decision-making, imagination, fictionality, and psychology; but it is not the only option for representation. The current paper investigates how using a model-based representation changes our understanding of immersion in pretense, engaging with fictional worlds, and acting in everyday life. To use this representation, 'belief' and 'desire' must be reinterpreted and certain derivative notions (liking, wanting, expecting, imagining) must be examined under this alternative interpretation. This paper is intended to provide a foundational account of the alternative representation technique via definitions, descriptions, and examples. Additional remarks are made towards identifying features of cognition that the classical approach has difficulty capturing but that can be included with the model-based representation and what benefits that provides us for addressing issues in pretense, fictionality, and ordinary behavior.

1 INTRODUCTION

There is a puzzle in aesthetic philosophy regarding why and how imaginative and fictional scenarios generate emotional reactions and actions just as real scenarios do, though often less vivid and with less force. It seems quite normal that we should feel genuinely saddened by the death of Anna Karenina, and yet simultaneously puzzling that events involving people who never existed should affect us at all. Reading about a devastating earthquake in Turkey in the newspaper will move many people to provide assistance of some kind, but reading that same

account in a novel will not motivate people to act in the same way.[1] The difference in action persists even when the emotional reaction is the same (and other features are fixed as well), leading many to posit the need for a distinct cognitive attitude for imagination (Schroeder 2006). Rather than adding another cognitive epicycle to an ill-suited decision/action theory I recommend explaining these phenomena with a more viable framework.

The bulk of literature exploring belief, desire, motivation, purposeful action, emotional affect, and other features of action, imagination, and pretense is still explored and explained via psychological models represented with a belief-desire propositional framework. Those psychological models, however, are woefully outdated and consistently contradicted by psychological, neurological, and behavioral research.[2] The idea presented here is that by using a model-based representation we can explain our behavior as generated by our *working models* rather than desires and beliefs. Working models (the occurrent system of objects, relations, behaviors, and associations being cognized) simultaneously employ folk models, scientific models, linguistic models, models of social interactions, models of human behavior, models of others' minds, and more. The purpose of this paper is to sketch a version of what a model-based framework would look like and outline how it can be applied to some of the questions regarding perception, imagination, action, immersion in pretense, and cognition more generally. I aim to provide a framework that we can use to construct explanations for mental cognition and in turn help explain our behavioral and psychological responses to facts and fictions, and how we use both in all we do.

2 TOWARDS A NEW MENTAL FRAMEWORK

The scientific evidence that our conscious experiences are merely epiphenomenal, and that mental events cannot have any causal efficacy, is nearly beyond doubt (Crick 1998, Wegner 2002). The philosophical support for the scientific claims is even stronger; the supervenience of

[1] Such fictional, hypothetical scenarios may prompt people to take precautionary measures to prevent or reduce the effects of possible future actual occurrences or other appropriate actions. The degree to which people react to such fictions likely depends on many features, but the point is that a person's behavioral reactions to a fictional account *is different from* the same account if it were thought to be real.

[2] Some newer research (Wegner 2002, Prinz 2004) lends evidence that insofar as behavior is motivated by cognitive processes it is produced through largely hardwired clusters of perception-manipulating and behavior-generating neurons. Much of what we actually do and what we consciously experience is filled in by cognitive mechanisms outside our control. This cognition is sometimes called *tacit knowledge* (or unconscious) background material or automatic (though possibly learned) response. One message from this research is that we may be ill-advised to model cognition based on what we consciously experience since the actual mechanism may be significantly different.

the mind on the brain (and the brain on chemistry and chemistry on physics) ensures that a better causal story is always available at a lower level of organization (Kim 1984, Bramson 2007). For example, the economic laws of supply and demand are in a similar position to mental events in this respect. Most would agree that there are no causal forces in social aggregates (and it is even part of economic doctrine that all purchasing behavior is generated by individuals). Yet supply and demand do seem to comprise an economic explanation of adaptive pricing phenomena. Descriptions of the same changes in the world at other levels of organization (e.g. a physical description of the phenomena in terms of molecules and atomic and chemical theory) may better approximate a causal story, but they will fail to be an *economic* explanation of the phenomenon. Similarly, the neurological description of some behavior may produce better predictions and have stronger claims to "truth", "validity" or "correspondence to the real world", yet a mental-level description in terms of our ordinary psychological concepts still seems useful and nontrivially explanatory.

2.1 MENTAL REPLACEMENT, NOT ELIMINATION

One desiderata of a mental framework is that it is *compatible* with neurological research. We desire to abstract from the neurological to the mental for the same reason we would choose to abstract from the atomic to the neurological (or economic) level: the loss in detail is traded for the ability to make different qualitative (phenomenological) distinctions. There are two obvious ways to achieve this compatibility: The first is to bin the neurons responsible for a particular mental function, and determine the relations among the binned neural clusters. This seems to be what boxologists think they are doing (Nichols 2003, Egan 2007), but by starting from introspectively available psychological states and going down they are getting the boxes all wrong from a neurologically informed construction. When we look for neurons grouped into their functional roles we will not, as a matter of contingent psychological fact, find regions for belief (or feelings of beliefs), desire, motivation, emotional affect, and the other categories currently deployed in descriptions of mental activity. There is no shortage of evidence for this conclusion (Kandel 2000 to start, but there has been a great deal of very recent research to overturn common notions that specific parts of the brain had these specific detailed functions; brain function seems to be much more integrated and dispersed than thought even five years ago).

The bottoms-up approach, exploring neural function and determining the produced mental states and/or behaviors, has a strong constructivist flavor to it that seems to ignore the

complex relations that apparently drive brain function and its phenomenal byproducts. Furthermore, the capability to measure and discern neural function is not yet up to the task: the state of the art is not remotely close to the necessary scope or resolution of neural analysis. Research indicates that such a binning of neurons by function is anyway impossible since neuron functioning does not organize itself into discernable categories. The bottoms-up approach of creating correspondences between neural operation and mental function also fails to inform our mental picture adequately.

A second obvious way to ensure compatibility of mental function with brain function is to make them explanatorily independent.[3] We will not attempt to find the brain functions that correspond to mental events; or even claim that there must be some specific ones. Instead, I propose to create a framework that is more compatible with neural research by being a more complete abstraction. While there is physiological and psychological evidence that the brain operates via modules of some kinds and that these may be responsible for particular models used here as explanatory tools, this seeming evidence is actually off the point of the current paper. So I do not mean to eliminate the need for mental descriptions and explanations by providing neurologically-based ones, but rather to replace the current neurologically and psychologically problematic propositional mental picture with a neurologically and psychologically compatible mental one. Simply stated, I aim to create a better match at the level of cognition and the scientific research comes into play only insofar as it informs our understanding of what cognition does and how it is experienced.

2.2 Against Belief-Desire Propositional Representation

To support the claim that this framework significantly departs from the standard belief-desire propositional (BDP) framework and provides benefits, some comments on what I take to constitute the BDP framework will be provided. This will not be a thorough review of existing literature in the area, but rather a sketch of the whole class of such constructions. Since my objection is a wholesale incompatibility with neuro-psychological research and an inadequacy to explain the mental and behavioral phenomena under study it will be sufficient to undercut those

[3] I do not propose to make them causally independent. The mind still supervenes on the brain, but we won't try to use descriptions of the brain to support our claims of mental function. The propositionalist can use this trick too; saying that however the brain works it produces sensations of propositions, probabilities, expectations, deductions, etc. That's fine. What it curtails is attempting to look at brain behavior and say "Here's where the propositions get formed" or "Here's where physics models get modified" or other such mind to brain claims. This would not weaken *my* case against the propostion-based framework.

aspects that any proposition-driven framework requires. To be sure some propositional models will be capable of accommodating some of the objections I raise, but I will demonstrate that none can accommodate all of them. Furthermore, the use of propositions to attempt to model such things, even when sufficiently successful, is kludgy and inelegant in ways that model-based frameworks are not.

The inadequacies (or at least extreme difficulties) of the propositional framework to accommodate emotional affect, urges, habit, pretense, psychological disorder, morality, stress, and a multitude of other attitudes and visceral factors is so well documented and established that it has become cliché to even mention that *homo sapiens* is a far cry from *homo economicus*. The contention that the model, though flawed, is still useful and appropriate for some specialized domains may very well be true, but those specialized domains are quite far from everyday experience and from immersion in pretense. The ever-growing collection of counter-evidence that people act according to the doctrine of rationality need not even be outlined here. I consider it now the status quo position that belief-desired proposition-based rational choice-like models are woefully in need of replacement and their continued use is only due to the fact that nobody has proposed a viable alternative.

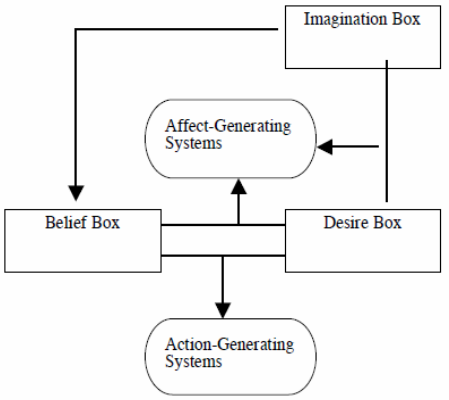
One defining characteristic of a BDP-based construction is that the currency of the decision mechanism is a set of propositions (statements that can be either true or false or some probability of being true). Desires, preferences, wants, likes, etc. usually come in the form of a ranking over the states of the world which are also described as ceteris paribus propositions (e.g. "I prefer that I have $20 to that I have $10." or "I prefer that 'I have $20' be true to 'I have $10' be true." or even "I prefer that 'I have $20' be at least half as likely as 'I have $10' than otherwise."). The required action theory in most cases is just "realize the state of affairs that I prefer the most".[4] In the standard decision theoretic approach emotions, urges, instincts, and other visceral factors are included either as 1) *variables* alongside the beliefs or 2) *parameters* for preferences and/or the credence of beliefs. Other techniques push some of these effects to the action theory and leave the decision theory unadulterated.

2.2.1 Boxological Generalization of the BDP Framework

Some BDP approaches are concerned with the form and paths of decision making rather than predicting or prescribing behavior and the boxological approach is one example of a more

[4] Action theories sometimes include many other features such as tie-breakers, direct action from certain perceptions (without needing beliefs about those perceptions), and many other features to fill in gaps between what is convenient to model as beliefs and what is necessary to produce reasonable reactions.

general approach. In this approach each category of distinct cognitive attitude is represented as a box with the causal linkages drawn with arrows (see diagram below of Nichols and Stich's boxology taken from Egan 2007). There is more to their mental picture than is included in the diagram, but the point is that it is perfectly consistent (and intended to be perfectly consistent) with the propositional account of the decision theorist: imagination reacts with desires to create emotional affect (just like belief) but does not create action (unlike belief). It is also consistent with accounts very different from the standard decision theory model, but it does rely on beliefs and imaginings having propositional content and on desires operating on those propositional contents to generate behavior and emotional affect. One thing that it implicitly leaves open is that some other, completely separate, mechanisms can also feed into the affect-generating and action-generating modules. But it also excludes other mechanisms from interacting with beliefs, imaginings, and desires to produce emotions and actions – though it doesn't explicitly preclude this possibility the authors do not mention these possibilities as relevant to the story. It is rather clear that these authors (and others) have a BPD framework in mind.



2.2.2 Coherentist Justification and The Web of Belief

Another variation on the propositional framework is a coherentist theory of knowledge (Bonjour 1985, Brink 1989). Coherentism is a reaction to problems in justifying beliefs using a foundationalist approach while simultaneously avoiding skepticism. Instead of there being a few select beliefs which act as a foundation for other beliefs, the beliefs form a holistic web of justificatory support with no beliefs receiving preferential status. As a simplistic example consider a belief that some perceived object is a functioning automobile. That belief is partially informed/supported by seeing it move, hearing it run, and identifying its main parts (wheels, windshield, doors, etc.). If, however, one saw that the tires were missing one would come to doubt that it really is a functioning automobile which would simultaneously affect ones beliefs regarding the engine's, brakes', transmission's etc. status. The presence of tires has no direct affect on the status of the engine, but one's beliefs are intrinsically tied together. I think that such webs of interconnected beliefs are an improvement on the linear, independent epistemic picture of the classical decision theorist, but the coherentist is also limited to propositional

content of mental states. Below I will revisit the holistic coherence idea in terms of model-based representations, but for now it is enough to note that coherentism is a step in the right direction, and might significantly improve the boxologist's case, but does not take us all the way to where we need to go.

2.2.3 Example of Inappropriate BDP Application

One rejoinder that the BDP may attempt to deploy is the fact that *any* behavior can be described in terms of (for example) the person having a set of beliefs about the world, having a preference ranking over the believed states of the world (contingent on one's actions), and a motivation to maximize preference. I claim that this description cannot provide explanation. Evidence contrary to the BPD position is readily available from everyday experience; it suffices to use the common "jerk my hand away from the flame" example. When I inadvertently put my hand into a flame I immediately jerk my hand out of the flame (and often scream). I could describe my behavior as coming to believe that my hand was extremely hot, in danger of injury, and probably easily removed from harm's way through various arm motions. I then access my preferences over worlds where I have burned off one hand and another where I have to recover from very minor injuries and all the worlds (levels of injury) in between…also accounting for the undesirable effort I have to exert to move my hand, the embarrassment of being caught in such a situation, and everything else. I calculate that my preferred world is the one with minimal injury and hence I am motivated to act to remove my hand from the flame in as little time as possible. That story can be told in perfect agreement with the BPD framework, but I can't imagine anybody thinking that it is a viable *explanation* for that behavior. The BPD descriptions may constitute appropriate explanations in certain restricted domains (e.g. corporate by-outs and FCC bandwidth auctions), but it certainly is not appropriate for general purpose behavior explanation: immersion in pretense, responses to fiction, scientific theorizing, and everyday life count among the inappropriate domains for BPD thinking.

2.2.4 Look, Propositional Representations Just Don't Cut It

It is clear that our cognitive faculty makes more detailed, accurate, and contextually contingent assessments of our environment than our linguistic ability to recall and describe our cognition can capture. This problem is vividly known to those who construct expert computer systems (e.g. automated medical diagnosis programs, or suspicious activity monitoring software); subject matter experts famously have extreme difficulty in accurately and effectively conveying

their expertise as true/false statements and conditional rules.[5] While this propositional language may be the only thing we have readily available to describe our cognitive processes, it is clearly insufficient to capture the cognitive processes that we actually undergo. It is not enough to posit that much of our mental apparatus is unconscious if one insists that it has the same form as our conscious thinking. So much more is going on, and must be going on, in our minds than what can possibly be represented by systems of independent beliefs, desires, motivations, motor functions, etc. to constitute our mental lives.

3 MODEL-BASED FRAMEWORK INTRODUCED

The contribution that this paper purports to makes is to provide an alternative to the ubiquitous proposition-based framework. Because I believe that the standard belief-desire apparatus has so thoroughly saturated the thinking of researchers in this area (and most other areas including our folk psychology) providing a lucid description of the new framework and a convincing exposition of its benefits is a far greater challenge than identifying the limitations of the entrenched propositional framework it displaces. I can only expect to produce a sketch of what such a model-based framework could be and since we can expect that the methodology will admit to numerous variations only a fraction of the conceptual space will be explored. My hope is that a combination of meticulous definitions, verbose descriptions, and eloquent examples will at least convey a rough understanding of the approach I have in mind and convince the reader of its potential value.

3.1 WHAT IS A MODEL?

Because a tremendous variety of objects, programs, equations, diagrams, rules, stories, and other things don the name "model" I cannot rely on any intuition that you may have to fill in what I mean. But despite the breadth of what one could consider a model, there are underlying commonalities: they are *representations* of something else and they purposefully exclude (*abstract* away) aspects of what they represent. Some might also stipulate that models are formulated or built for a purpose or benefit. Supposedly the expected benefit that the abstraction produces by excluding elements from the representation will guide the choice of what aspects are abstracted

[5] I have references for this claim (they are easy to come by) and this difficulty really is famous within this branch of artificial intelligence. I will track the best example down for the next iteration of this paper.

away. But I don't think purposefulness is the correct thing to attribute; rather it is *functionality* that models possess.[6] This is also an important distinction when we realize that the mind automatically (and therefore not purposefully) generates models which function to reduce the number of entities and relations perceived and make cognition feasible (a significant benefit). Let's have an example of this mental abstraction.

3.1.1 Senses and Perception

When the mind perceives (say) an apple from the smattering of hues and values in one's visual field it has abstracted from the sensual micro-data to the level of everyday object mental representation. Furthermore, the current scientific consensus is that there really are no apples in the world to perceive at all; objects at that scale are composed of atoms which in turn are composed of ever-changing sets of quarks and leptons. The apple concept itself is an abstraction in the sense that it excludes all the details about the subatomic particles that make it up. But we (and other animals) didn't have to learn that apples are made of tiny particles to perceive the apples when we were actually seeing particles. We just naturally perceive the apple as an object in the world. To some degree the level of perception is determined by our size and the laws of optics, but there is a strong and obvious argument to make that perceiving the objects we do perceive the way we do perceive them conveys an evolutionary fitness benefit.[7] For many intents and purposes (e.g. eating, purchasing, juggling, cutting) considering the apple as an object makes sense, but if we are considering whole apple trees, orchards, or countries then thinking in terms of apples (and other apple-sized pieces) is overly burdensome; we just can't

[6] Not all abstractions, not even all representative ones, are models. If all abstractions were purposeful then all models would be, so I must be claiming that some abstractions are not purposeful (but may be functional). If this seems counterintuitive then the following example might help elucidate what I mean. Place a magnet on a tabletop and a piece of cardstock over it. Then sprinkle some iron filings over the cardstock and watch them scatter to align themselves with the magnetic field lines. The pattern of iron filings functions as a visible manifestation of a slice of the magnetic field produced by the magnet. The arrangement of filings produced by the field is not purposeful, but it still performs a functional role (i.e. representing one system in terms of another). One may try to object that the setup to use iron filings to function in this way was purposeful, but this is a false start since the same thing would happen if a crate of iron filing byproducts from a bolt and screw factory accidentally crashed open near an industrial magnet: we would (accidentally) have created a model of the magnet fields with iron filings. Radioactivity, cathode-ray tubes, Earth's magnetic field changes, and several other phenomena were discovered accidentally because certain systems naturally function as models of other systems.

[7] One extreme example of the fitness benefit determining our perception is the idea that the universe might actually exist in four or five or more spatial dimensions but that we only perceive in three spatial dimensions because that was sufficient for our proto-human ancestors to avoid predator attacks and so perceiving in other dimensions, if it ever evolved, simply didn't persist because it offered no fitness advantage. A more mundane example is the differences among animals to perceive in different color ranges. If humans were able to see in the infrared or ultraviolet spectrums as some insects and fish can then we might differentiate objects differently than we do.

keep that many things in mind at once. So we can abstract again to these other levels and we do so naturally and without consciously thinking about choosing a level of abstraction. When we're considering trees we see trees and naturally ignore all the apples on them (as individual objects) and when we're considering apples we naturally ignore the trees they hang from (as objects to be perceived).[8] The objects we perceive in the world and in our imagination constitute the ontology of our models. A similar automatic context-dependent perception holds for properties, relations, behaviors, etc.

3.1.2 Cognition Requires Parallel Models

The mind can digest many kinds of formal models cached out in terms of equations, diagrams, words, and other symbolic/formal representations but (I claim) an internalized mental model has a form that most closely resembles mereological models: models specifically constructed to represent and explore part-whole and part-part relationships. These can be best *described* as mental holistic simulations of how things play out, but I do not think that cognizing them is necessarily similar to any sensory experience. The most intuitive analogy is to a visual playing out of a scenario. One reason that the visual analogy is good is because it is one of the few senses with which most people can take in a whole scene and individuate its parts.[9] Another reason is that though our verbal description of any scene must be done in serial fashion (describing one part of the scene at a time), in an actual scene many things happen in parallel (and all at the same time). The brain and our cognitive functions work in parallel fashion (Kandel 2000). We may be limited to a one-track consciousness and linguistic ability, but these are only surface-level cognitive functions.

3.1.3 Components of Models

A model-based mental representation (as it is being used here) is a holistic rule-driven "simulation" of a scenario that typically consists of four parts: 1) an ontology, 2) a set of relations among the objects in the ontology, 3) an environment or space, and 4) dynamics. The ontology is comprised by objects and their properties. The relations can be of any kind – physical, social, causal, mereological, temporal, informational, etc. – and multiple types can be

[8] We might instead be interested in plant development, metabolism, and reproduction or the chemical basis of nutrition and decay or other processes that require thinking of an apple as the whole and the elements of the model to be the relevant parts of apples.

[9] I believe that an auditory analogy is just as good, but most people have not developed the ability to (say) hear an orchestra and keep each instrument's contribution to the overall sound separate. And we rarely think of the instruments' parts as causally related, which will become important in mental models.

included simultaneously. Insofar as some models are about objects "in the world" the objects will be embedded in a physical space, but not all spaces are physical and not all objects and relations need a space. Model dynamics include changes in the set of objects in the model, changes in object properties, relations, spatial orientation or location, and (of course) changes in how things change.

To help focus on what I intend to be talking about you can consider the example of teaching a classroom of students. There would be student objects each with their observable physical properties as well as other properties that one may know from previous interaction. You may also be aware of various social relations among the students such two being roommates, some being in another class together, etc. There would also be a teacher object that would be associated with the self. Being a teacher-typed object implies certain typical properties (age range, education level), relations to students, location in the classroom, and behavior. Being associated with the self, that object would also embody certain relevant properties that you yourself have (your name, whether you wear glasses, ability to walk). The teacher and the students would all be located in a classroom with a whiteboard, a chair, a table, several desks, windows, a door, etc. Different people with different background experiences would imagine different classroom scenarios given just the prompt "pretend you're a teacher", but that description should help clarify what these mereological models are like. I'll use some specific examples throughout while discussing the features of the framework in more detail.

3.1.4 Endogenous and Exogenous Elements

I will assume here that our mental models are finite. The size of a model – all the things in it, its space, and its extent in time – is called its *scope*. All elements inside the scope are endogenous aspects and so naturally all those features of a model that are not explicitly included in the model itself (or derived or emergent from elements in the model) are its exogenous aspects. Sometimes models are embedded in or linked to other models, so an exogenous aspect of one model may be endogenously determined by another model. Thus one's physical and moral models are (probably) separate modules, but they might be called upon to regulate aspects of a scene involving dropping an anvil on a person's head.

For an example of the context dependence of exogenous and endogenous aspects let's consider two physical models. The first is relevant for catching baseballs wherein the value of gravity is considered fixed and determined exogenously. A cosmologist may have an intricate model wherein the strength of gravity is an endogenous aspect, but even for that cosmologist the

strength of gravity would be exogenous for models involving everyday medium-sized dry goods. For most people the strength of gravity is always exogenous, rare are the people with models that generate gravitational forces and most people do just fine without them.[10] Another type of exogenous feature is the specification of what plays the role of the atomic (or baseline) elements of the models (i.e. its *resolution*). Even the scope can be determined endogenously (growing and shrinking systems), but what objects count as parts (the ontology and relations) cannot be endogenous. I previously claimed that part determination for our models (at least our folk models) is automatic in a sense, and thus determined exogenously by our sensory perceptions (and perhaps by cognitive structures too). Shifts in resolutions will then be just a special kind of model switching that preserves part-whole associations (more below).

4 FEATURES OF THE FRAMEWORK

In this section I will attempt to describe many of the features of cognitive models and bring them together to forge a clearer picture of what the structures are and how they work. Because there is some slippage in what is a model and what is a component of a model (and especially considering that models can be components of or linked to other models) being clear and concise is difficult. Within the space and time allowed I can only hope to provide a sketch of the framework sufficient to ground the examples and benefits of thinking about cognition in terms of mental models.

I differentiate one's *working models* from one's *vocabulary of models* by identifying the former with occurrent cognition and the latter with one's cognitive dispositions. The vocabulary includes all those models that contexts can prompt into the set of working models and the working models include all those cognitive structures that one's current prompts (sensory or "dreamlike"[11]) have activated. The working models need not be consciously operating; conscious/unconscious is not the distinction being made here. If somebody has not eaten in a long time then their *seek food* model would become working (which makes them irritable, more sensitive to scents, salivate in the presence of food, etc.) though that person may very well be

[10] There will be more later on gravity and what constitutes our understanding of gravity in these everyday applications where there clearly is no explicit role for gravity in the mental model.

[11] I use the word 'dreamlike' to refer to perceptions that do not come from sensory input. These would include perceptions from dreams, daydreams, hallucinations, and any mere positing of an entity. There is a strong inclination to refer to them as purely imaginative perceptions, but that would mean something different in light of how I use 'imagination' below. Perceptions from fictional accounts (movies, books, etc.) are sensory in this sense.

unaware that hunger was driving these behaviors. Once the person has eaten something the *seek food* model leaves the set of working models but is posed to return when it receives the appropriate prompt from the body or other model.

For the concepts and processes described below it is difficult to explain, and perhaps to understand, my meaning because it is (I think) significantly different from how we usually talk and consciously think about such things…but introducing an alternative to the ordinary belief concept is also a point of the new framework. An example of a difficulty is that I generally will be referring to representations of objects themselves rather than propositions about objects, yet the only language we have to refer to things is using propositions – this issue makes some of the language convoluted and the ideas difficult to tease apart. And please be patient in reading through this section because each subsection is interdependent on the others, but must be presented in some order or other. Some passages are also long and dense and may require rereading. I hope you will find the effort to be worthwhile.

4.1 CREDENCE AS AN ALTERNATIVE TO BELIEF

The concept of *credence* in the existence of an object, property, relation, position, or dynamic rule is close to the concept of belief. And by 'belief' in the previous sentence and everywhere in this paper I mean the propositional sort usually utilized in philosophy and decision theory. Credence may be thought of as the strength of the feeling that the perception belongs where it is within the mental model. Or it may be considered to be how much the object exists. It can also be how well 'X bears the A relation to Y' matches the world (fictional or real). Everything in every model has a credence level, and most will have separate credences for different modules[12], different contexts, and contingent upon other aspects of cognition. It is **not** how much one believes that the element is there – credence is *how much the element is there* within the model. Credence of an object is nonpropositional in the same way that the existence of dogs, triangles, concertos, kindness, etc. is nonpropositional. And, for example, one does not have a degree of belief that a fox is a dog, but a fox partakes of dog-ness to some degree. So if one thinks of credence of associations as how well concepts "hang together" then the credence of things themselves can be thought of as how well (or to what degree) the thing's concept agrees with the 'existence' concept.

[12] I am using 'module' to refer to a model that has densely and strongly connected internal associations and a few specific "input" and "output" associations. Wherever 'module' appears, 'model' can be replaced.

At first glance the notion of credence here is somewhat similar to coherentist notions of truth and belief (Brink 1989 and §2.2.2). Coherentist theories are expressed as networks of propositions such that the truth of each relates to the truth of others held in a holistic way. This relation is characterized as requiring certain propositions to be true if certain others are true and/or false; interrelations specify restrictions on which propositions can and cannot be true simultaneously.[13] Insofar as models form a set of interrelated objects, properties, interactions, dynamics, etc. and that the credence of some of these can influence the credence of others (through inference, reinforcement, expectation, confirmation, and many other feedback mechanisms), there are more similarities to coherentism than to standard propositional belief. But recall that coherentism is cashed out in terms of the truth and falsity (or degree of truth) of the connected propositions rather than in the degree of existence of objects.[14]

Another point on which credence differs drastically with propositional belief is that degrees of belief are generally given a probabilistic interpretation: a .8 belief that *p* means that in four out of five cases *p* is (or is expected to be) true. No such probabilistic meaning can be given to credence (though see 'credulity' below). Let's say I am considering Paris. There will be an object in the model I construct for the Eiffel Tower and it will have a very high credence because I have received many inputs indicating that the Eiffel Tower is in Paris. Given the shape and size of the structure I also attach the property 'made of steel' to the Eiffel tower, but with a low credence because I have not actually heard that it is made of steel. I might also attach 'made or iron' with a low credence and 'made of selenium' with an extremely small credence. My model for such things indicates that it can be made of only one of these metals, but given what I know the three properties can all obtain with low credence. This does **not** imply that I believe it is made of all three metals or that I assign some probability to it being made of each of these metals. In the model-based framework it just means that when I consider the Eiffel Tower these other 'made of ____' concepts are also included in my working model (i.e. brought to mind) at varying degrees of intensity.

---

[13] The belief operations of a Pearl Graph or Bayes Net and any of the other similar constructs are strongly coherentist. The belief of the whole depends upon the truth values (confidence levels) of the inputs and the structure of the connections. In practice these tools usually take the forms of trees (which would make a very boring coherentist picture), but if built with feedback they could model something with a coherentist picture. As will be made clearer later the current models are therefore not representable with coherentist techniques because they are not probabilistic or propositional. This is a tangential technical note and can be ignored if it falls outside of the reader's expertise since it is describing what a model **is not**, and we are more concerned with what a model **is**.

[14] There is a slippage here to avoid. The models used here include representations of objects themselves and the credences form a network for the strength with which the representations obtain. The coherentist approach instead has beliefs (possibly about the existence of objects) and interrelations regulating the value of the beliefs.

4.1.1 Credence as an Alternative to Belief-Like Imagination

The question of whether imagination is a distinct cognitive attitude from belief (Nichols 2003, Schroeder 2006, Egan 2007) need not even be asked because those categories need not be used at all. Within models the credence carried by the elements is the same concept regardless of whether a working model was prompted by our senses or by "dreams"[15] and whether it is about a real-world scenario or a fictional one. The question of how exactly it is that some models produce physical action (and others don't) demands attention. It could be that the difference is noncognitive (Prinz 2004) and the body's motor neurons take cues from their physically excitatory state to generate physical responses. It could be that there exists an implicit mental model that generates physical behavior based on features of the working model – and frequently "gets it wrong".[16] But note that imaginative models are sometimes *meant* to generate behavior; that is what make-believe playing is all about. I take the mistakes and play to be strong evidence that "real-world belief" (credence in models prompted by sensory stimuli) and "imaginative belief" (credence in models not prompted by sensory stimuli) are **not** different, but that some independent mechanism (imperfectly) regulates which working models generate behavior and which do not (more later).

The similarity of the cognitive attitudes becomes more obvious when we recognize that our working models (whether real-world or fictional or make-believe) are necessarily populated with a great deal of material from our model vocabulary. Some of that (background) material may have been learned from sensory inputs in the past, but as they are used the structure of one's cognitive models are not populated solely from what is sensed. One constructs an intricate, dynamic web of how things work in different situations (the model vocabulary) and applies modules from that material to fill in details in the working model (which is generated from occurrent prompts). Pulling out modules from the model vocabulary and incorporating them into the current working model is a kind of imaginative act that is required for any kind of cognition. This idea is explored in more detail in *A New Role for 'Imagination'* below.

---

[15] Recall that I am using 'dreams' to mean all those thoughts that are not prompted by external stimuli. These are things that we "just think up out of nowhere" or so it seems.

[16] The thought-behavior adaptor "gets it wrong" whenever it fails to produce actual behavior when that is appropriate (e.g. paralyzed with fear, in a daze) or whenever it produces inappropriate actual physical behavior in response to imagined or fictional situations (e.g. screaming awake from a dream, flinching from movie violence).

4.2 PROCEDURAL KNOWLEDGE AND ITS KIND

I also implicate *procedural knowledge* as a relevant alternative to standard propositional belief. Procedural knowledge is typically described in terms of skills, habits, accumulated dispositions, biases, and other behavior that people learn by doing. I know how to ride a bicycle, play a violin, and punch through a board – and I had to practice all these things to come to be able to do them – but they are not stored and processed by my cognition as intricate sets of beliefs. Though it is the case that I believe that I can ride a bike, for each required component of that skill I do not have a belief about it despite having unfaltering working knowledge to do it. "choose bright-colored tree fruits instead of dull ones" might accurately describe a rule (bias) that my behavior follows even though I can be unaware of such a consistent preference.

Natural tendencies have a similar representation and often must be practiced to develop proficiency. Walking, for example, must be learned in some sense but we don't have to be told to try to do. Perceiving a huge smear of colors and shades and being able to identify objects, properties, three-dimensional relations among them, etc. is beyond the abilities of newborns and common among adults. This ability does not come about by acquiring beliefs regarding how colors and shades relate to objects and relations. This just happens.[17] When we see a vase about to tip over we don't recall beliefs about gravity, the fragility of glass, the hardness of marble floors, and use them to produce beliefs that the vase will soon be a heap of glass shards on the floor unless we intervene. It is not the case that we act because of that belief and a desire to keep the vase from breaking. It's more like riding a bicycle. After the fact we can assuredly come up with a collection of beliefs and preferences that would (if held) motivate the action and describe the scene. But the availability of such a propositional description lends no evidence to a claim that we cognitively hold such a description or that a description in those terms has any explanatory value.

Actions like riding bicycles and scenes like catching tipping vases do prompt a response and there is some indication that particular parts of the brain bear the representational burden.

[17] Neuroscientists have recently put in a great deal of effort to figure out how all this happens, but the neurological details are largely irrelevant to the mental model. The features I describe here and throughout are based on the standard "textbook" neurological picture from Kandel et al's *Principles of Neural Science* (Kandel 2000) in addition to other papers read over the years and not referenced for lack of space, time, and relevance to the main point. One work worthy of special note is Jeannerod's *The Cognitive Neuroscience of Action* which uncharacteristically includes motor-oriented representations. The bias towards only being concerned with object representation (rather than performance representation, motion representation, rule representation, etc.) is common in both the neurological and psychological research and has, so far as I have seen, extended into philosophical work.

But whatever it might be that neuronal patterns store and reproduce it is clearly not propositions. Propositions could describe it (or rather we could describe it with propositions) but *how to do things* or *how things play out* or *how to react* are not natively represented propositionally. Whether the vase actually falls and breaks or not will depend on how well-matched one's model is to reality in this instance (see *sense-confirmation* below). But even if one expects it to fall, lunges at it, misses the vase, and it bounces because it is actually made from plastic, nothing in that scenario is false.[18]

4.2.1 Gravity Example

Procedural knowledge of this kind is more far reaching then one might at first suspect. Catching a baseball, hitting a target with a golf ball, swinging from one branch to another, applying the correct amount of force to climb exactly one stair, are all example problems that to solve with physics equations requires a value for the force of gravity. When using Newtonian equations we must have the requisite belief that gravity pulls at 9.8 m/s^2 (or some other value) to solve the problem and determine the required amount of exertion. But it would be ridiculous to think that people (not to mention squirrels and grasshoppers) are solving these equations to perform these actions. Creatures have been jumping from branch to branch since long before anybody had a belief about the force of gravity on Earth. Some might claim that there is an implicitly held belief; I would counter that describing their actions in language requires attributing to the actor a belief of the strength of gravity and it must be an implicit one (especially in the case of non-linguistic animals). But that is just a description – and a more accurate description would be that the agent's behavior is consistent with a belief in a certain value for gravity, certain other beliefs about the effects of certain levels of exertion, beliefs about wind friction, etc. Attributing beliefs to creatures might make talking about them easier, but those beliefs are not in the creature.[19]

Our understanding of gravity (in the model-based framework) is fully captured by how objects behave in the relevant contexts. Recall that many models are akin to mental simulations. Given a particular scene (either prompted by our senses or dreamt up) the objects' physical orientations elicit concepts regarding their behavior. Objects in the scene suspended in mid air

[18] Nothing is true either. Truth and falsity do not apply to these scenarios or procedural-like knowledge. Other evaluative concepts do apply – such as more appropriate, more efficient, and more consistent – but only content-bearing propositions seem capable of delivering truth value.

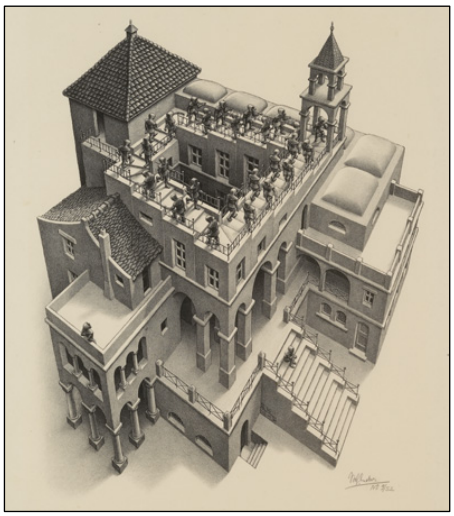
[19] An even clearer example is the attribution of beliefs to computers to describe their behavior: "My computer believes that it's in the Pacific time zone." The description is consistent with the behavior but it does not explain the behavior in a way that is kosher for computers…or (I claim) for people either.

(for example) will link to our gravity module which produces expected behavior in accordance with how things (typically) behave when they are not supported (i.e. they fall). *Our understanding of gravity really just is the behavior of objects expected in the appropriate contexts.* People whose expectations match observed behavior have a better understanding of gravity. It is similar to procedural knowledge in that gravity just does what it does in the model – my body just does what it does when playing the violin – and that operation need not access anything resembling a belief, desire, affect-generating systems, heuristic, node association, etc. I claim it can be *better represented* as contingent behavioral rules in mereological models than with beliefs and desires, but I am not claiming that such understandings exist as such in the mind. I claim that most such things *just happen* in the mind, but they can be represented as elements in mereological models.

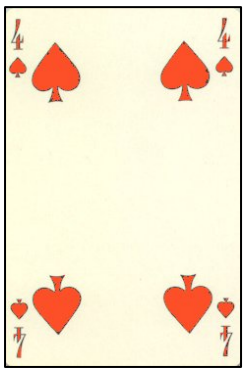
4.3 Perception and Model Building

For our mental models, the baseline ontology is provided for us by perception. In some sense we get it for free. Neuro-psychologists have a very difficult story to tell about how our visual field, auditory patterns, smells, tastes, and touches become experiences of objects, relations, etc. There is a strict sense in which one never actually senses bicycles, one senses patterns of color and recognizes a bicycle (assuming the viewer knows what a bicycle is). Usually one can perceive unknown objects, have a stand-in object in a model, and then learn about it through expectation, experiment, and updating. This is how we can learn about new things in our environment.

However, if the object is different enough from our experience and contrary to our expectations we can actually fail to perceive an object or misperceive relations. Optical illusions sometimes exploit incongruence with normalcy to produce extraordinary (and therefore interesting and scintillating) results. One example is Escher's famous infinite staircase; another is the Penrose triangle. These illusions utilize the fact that our mind interprets perspective from several reliable cues from the environment and creates a model that contains spatial relations of the perceived parts. These generally reliable perspective cues can be combined to create an impossible combination of relations: each



corner is admissible to a model, but the combination of all four is not. Also recall a famous (even if untrue) anecdote about how natives in the Americas were unable to perceive Columbus' ships (supposedly because they were so different from anything they experienced) even though they could see the wake they created and knew *something* must be out there. Eventually they got it, but it required a more direct perception/interaction to learn what they were and make it congruent with their models of what exists and how they work. Recall also the Bruner and Postman experiment of showing participants anomalous cards such as red spades and black diamonds; the standard deck's characteristics were so thoroughly ingrained that participants had delayed responses, perceived things differently than they were, some suffered severe *cognitive dissonance*, and some even called into question their own knowledge of what a spade *is* (Bruner 1949). There are plenty of other mundane examples of these phenomena, especially in education and developmental psychology. The main point here is that sensory input often must link (be associated) with a concept in a model for it to generate a perception.



In our perceptual "business as usual" observations of objects, behaviors, relations, etc. generate corresponding objects, behaviors, relations, etc. in our working models. When this is done the perception is matched to concepts related to it from previous experience and at various levels of abstraction.[20] The credence with which properties/associations[21] adhere to the object depends on how well perceptions match with the concepts to which the properties apply. A similar thing happens for perceptions of things other than objects: actions, patterns of behavior, linguistic information (see below), social rules, etc.

4.3.1 Possible Variations in Perception

A particular scene may generate a perception that is interpreted as representing an act of kindness; such a perception would occur in the interaction of at least two objects (people or maybe animals) and would have a property of kindness, a degree of the kindness property, a

[20] As before, the neurological story of how this happens is not fully understood and need not be known to justify this claim. That perceiving (or recalling) some object, behavior, or raw feel can and often does conjure other associated concepts is familiar to most and widely accepted. Furthermore, this associational aspect seems to apply to memory linkages, stimulus/response reactions, sets of properties (e.g. colors and flavors), and so many cognitive functions that it *probably* reveals something about the structure and/or operation of brain.

[21] I mean to use 'association' in the broadest possible sense to include the standard properties of color, mass, motion, temperature, etc. as well as relational properties (e.g. above, left of, inside), dispositional properties (e.g. fragile, honest, flammable), dynamic properties (e.g. robust, stable, path dependent), mereological properties (e.g. a part of X, has parts {a,b,c}), and any other sort of concept that attaches to other concepts including models and the outputs of models. Some properties have properties or other associations themselves.

direction, and those sorts of things. Perceiving kindness might require some kind of training and may be a natural capability (like walking) or it may be a cultural one (like doing calculus) or some mix (like riding a bicycle); though it might very well be an innate direct perception (like perceiving redness or warmth). Or kindness may turn out to be an abstract property that supervenes on other perceptions (e.g. two creatures being brothers) – in this case associating kindness to the interaction will require importing the necessary model(s). All these variations count as perceptions, and taking something with ambiguous sources (kindness) as an example highlights the fact that the details of how a perception figures into our mental models may not be clear despite the fact that its perception is clear. Not knowing how we perceive kindness does not interfere with our perception of it.

4.3.2 Perceptions from Linguistic Material

Written and spoken language, as well as signs, symbols, gestures, and other forms of linguistic communication, are sensed in the same way that other objects, relations, properties, etc. are sensed: through sight, sound, smell, taste, and touch. In these cases, however, it is doubtful that any natural brain function produces the appropriate perceptions. Converting linguistic material into perceptions certainly requires models (given the model-based representation used here). These linguistic models associate certain sense data with concepts; e.g. the sound /kæt/ with small furry purring domesticated omnivorous felines that lick themselves clean and shred furniture as a greeting. More than just a cat object is perceived, the whole relevant module or model segment is activated (added to the working model) in association with the visual perception of 'cat' or from hearing /kæt/ or seeing the ASL sign for 'cat', etc. My positing that there is a linguistic module that associates sense data with mental concepts (i.e. elements of models) is rather close to a boxological view of how such things might work; independent modules are like boxes. But by describing them as models I can bring in the material on how models are initially generated, activated in one's working model, and utilize expectation, liking, desiring, imagination, and all the other aspects of models to explain how language does what it does and how it interacts with the rest of our cognitive abilities. As you might have guessed, doing all that would take us too far from our focus and would require more space than is available in the current work.

4.4 EXPECTATION

At the most general level, *expectation* is the strength with which one feels the elements of models involving scenarios that have not yet happened. This obviously involves temporal considerations, but models already include dynamics and dynamics affect (among other things) the credulity of the elements in the model over (mental, simulated) time. This definition allows us to not only identify how expected something is but to compare the level of expectation of two scenarios…even purely fictional scenarios.[22] Elements in models, regardless of the model, do the same sorts of things (reinforce credence, elicit other elements, bring in associated models, derive from and drive perception, etc.) and so expectation works the same general way regardless of what the model is of. At less general levels, significant variations in *expectation* arise.

It is easy to appreciate the difference between 1) an expectation formed upon hearing that a T-ray probe will fly by Mars next year and 2) an expectation formed upon hearing car tires screech. Perceptions that are far from our personal experience (e.g. abstract, remote from daily life, unnatural), far from us in time, or in any way alien to us will attach lightly to our working models (i.e. low credence). Remember that credence is not propositional belief and so low credence does not mean low probability; one can be completely convinced that the T-ray probe really will fly by Mars and still give it low credence because the relevant portions of one's models is only weekly experienced. Insofar as experience populates our models and generates associations, familiar scenarios will naturally possess a larger, tighter webs of reinforcement and thus greater credences overall.

I further claim that there is a significant difference of kind in different expectations, not just one of degrees. The differences in kind reveal themselves as different kinds of elements. Expecting an object to be in a box is a different sort of experience than expecting somebody to be another's brother which is again different from expecting an object to fall. And those are different than expecting that the force I apply to my violin bow will produce the appropriate volume and tonal quality. All of those different experiences get captured in models as credences in (different kinds and combinations of) elements of models for things that we have yet to

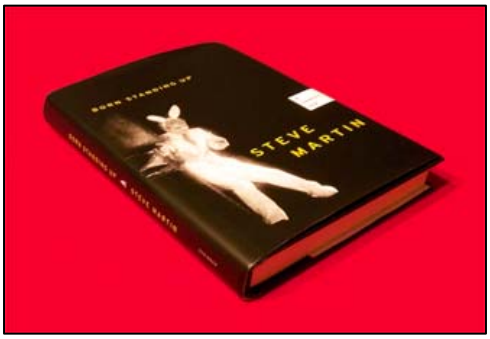
[22] Such a comparison will require some way to aggregate the credences of individual model elements into a level of whole-model (or model portion) credence. This might be thought of as something like an averaging of the element credences, but for the present purposes we need not have any particular formal procedure to aggregate the credences, only that somehow people can experience credences of model portions and (perhaps roughly) compare them.

experience.[23] But just because they are captured in models in similar ways it does not mean that the psychological experiences and behavioral responses are similar. I mention this here to highlight the point that this is a model-based framework for *cognition* rather than a story about experiencing mental phenomena.

4.5 A New Role for Imagination

Within the propositional framework, 'to imagine' was used to mean a fictional version of 'to believe'. And though 'imagination' has also been used for other things in this area of research, I have a very specific role for the term (that I also think is consistent with our intuitive notion of imagination). The imagined elements of a working model are those that are not perceived through our senses. These imaginary elements, it turns out, will populate a vast majority of our working models and also play an important role in converting sense data into perception.

When somebody sees a book on a table she only actually sees a 2 dimensional shape in various hues and values (see photo of book). The mind then interprets that via some mechanism as a book object in a model with some credence and possibly as other objects (e.g. a



cleverly disguised coin bank) with other credence values. I claim that the perception of that object as a book includes automatically "loading" features of books into the working model (e.g. hinged on one side, comprised of many thin sheets of paper mostly filled with words, light-weight, flammable, and sturdy enough to balance furniture on). To be a book it needs to be capable of opening, but one need not actually see the book open to perceive it as a book. So, I claim, the filling in of book-associated details with the sense-prompted perception of a book is a necessary role for imagination. And this role for imagination is necessary for the perception of **anything**, whether prompted by the senses or by "dreamlike" positing of entities.[24]

---

[23] Actually, the element(s) over which we experience an expectation may have occurred in the past and have been generated in the past. What will turn out to be the relevant feature is whether the elements were perceived via the senses or through imagination (next sub-section).

[24] As I have mentioned elsewhere, one may feel it natural to call these "dreamlike" prompts and the perceptions that accompany them "purely imaginary", but I am using 'imagination' to refer specifically to the filling in of details in the working model from associations to the model vocabulary and other elements in the working model. Thus "dreamlike" prompts are not themselves imaginative (even though that word seems to fit) and are rather the internally generated equivalent to sense data.

I'll say that again a little differently. In a working model, some elements are prompted by sense-perception, others are filled in by models, and still others are merely dreamt up "from nowhere". As one receives an account, hears a tale, sees a scenario unfold, etc. one constructs a working model wherein certain elements are prompted by the senses. Those elements in turn prompt (activate, bring in) other material from the model vocabulary which expand the working model. The expanded working model may contain new combinations of material that prompt yet more material from the model vocabulary (see *prompt cascade* in §5.3). All those elements would be imaginary in the model. Those imaginary elements of models form expectations of what sense-perception might deliver. When one sense-perceives an element that exists imaginarily we will call it sense-confirmed (i.e. prompting something that is already in the model), but that sense-confirmation may or may not increase the credence of the element. For example, actually seeing that book open will not likely alter the credence of there being a book there (but see *credulity* below), but it may strengthen the association that books are things that open. Sense-confirming (or sense-opposing) has the role of satisfying (or frustrating) ones expectations.

Imagination plays the major role in the feedback mechanism among sense-perception, experience-based models in the vocabulary, associations among models and elements, and expectation. One sense-perceives objects, relations, behaviors, etc. from the environment and builds models from these observations that are (over time) put into the model vocabulary. The elements in a model become associated with each other because experience has shown that they "hang" together. The future sense-perception of one element of a model will make one imagine the existence of all the associated elements in the model. Those imagined elements are precisely what one expects; those imagined elements are the details one fills in given what one senses and what one has learned about relevant situations. Further sense-perception will then strengthen or weaken the associations of the elements and/or their credence values in the relevant model portion depending on the congruence between imagination-driven element expectation and element sense-confirming.

Hence imagination, as it is defined here, is a necessary component of all cognition. When one perceives a tipping vase (regardless of whether it is a sense-prompted or dream-prompted perception) one imagines how the scene might unfold, including possibilities contingent on one's own actions, based on previous experiences with associated situations and generates expected scenarios for each possibility (in parallel). Such imagining can adjust credence (and thus expectation) levels of elements through prompt cascades and coherency

checks. In some situations the most strongly felt possible scenario will include action on the thinker's part and she may *just do* the action imagined/expected in the model (again, regardless of whether the scenario was sense-prompted or dream-prompted). However, sometimes the working model will include choice behavior over multiple possibilities and so I am going to spend a little space in the next section discussion model-based alternatives to propositional desire and motivation.

4.6 LIKING AND WANTING

As mentioned before, there is strong neuro-psychological evidence that our conscious experiences (especially of decision-making) are not causally efficacious for our cognitive process – the mental activity actually responsible for generating behavior (Kandel 2000, Wegner 2002). But that does not necessarily imply that there is no cognitive processing specific to decision situations (in contradistinction from impulses, stimulus-response, compulsions, breathing, and other behavior that may be cognitively regulated but not occasions for choice). Recent neuro-psychological research has uncovered distinct brain regions for cognizing *liking* and *wanting*; and on reflection these do seem to be distinct concepts.[25] Liking is an attitude one holds towards perceived elements in the working model, it is passive in the sense that one must already be perceiving the element to like it. Wanting is an attitude that one may have towards models in the model vocabulary though the wanting persists when those elements are activated. Wanting can be thought of as dispositional in the sense that one will like perceived elements that were wanted. In addition wanted models have an association with dreamlike prompts such that they come to mind (join the working model) seemingly on their own (a daydream fantasy).

I admittedly have not yet worked carefully through these concepts. My current thinking is that liking and wanting can be tied closely to expectation: the similarity in what is expected (imagined credence for perceived elements) to what is wanted (the credence of those elements in the imaginary model capturing the wanted world) determines how liked something is. If one does not have a relevant previous experience, and hence no model in the vocabulary and no

[25] The point of mentioning the neurological research is that since there are distinct brain regions that activate with very different stimulus it is likely that the two categories are processed differently. Since the BDP framework lumps them together as 'desire' it cannot accommodate these conceptually distinct cognitive attitudes. Adjustments could be made, e.g. giving each their own box in the functional map, but this would be a straightforwardly ad hoc addition (epicycle) to reconcile the theory with the evidence. I don't have the reference for this research at hand, but Peter Railton has it (and I have it back in Michigan) and I can get it if you are interested.

association to a prompt, the person will not have wants about that prompting element and will not experience liking or disliking that element. That is a start. It may be a push in the wrong direction, and at best much more needs to be considered to make these cognitive attitudes coherent with the rest of the framework. But liking and wanting are **not necessary** components of the framework. I can rely on a procedural-like function to generate behavior based on perception and expectation alone. For example, I may claim that differences in expectation and sense-perception generates *cognitive dissonance* which generates behavior to minimize those differences based on available associations with models (which, as you will recall, includes behavioral elements).

4.7 More on Credence and Introducing Credulity

Credulity is defined as a level of confidence and is a notion much more similar to belief than credence. Credence does most of the work in building models, generating plausible implications, and determining how clear our picture of the (fictional or real) world is. Credulities instead measure how certain we are that some model is appropriate, that some object does or does not exist, and how likely things are to happen. But even credulities need not form the standard probabilistic notion of belief. It will be more useful to utilize a concept that accommodates a distinction between *not believing* and *believing not*; the former is ignorance and the latter is a positive attitude that something isn't the case. Despite its variance with standard probability such a notion can be built into a mathematically sound and unified formal theory (e.g. the Dempster-Shafer method of evidence accrual). While that may be nice to know, it is more important for the current purpose to recognize that we need a difference between an object's being in existence in a model and our degree of confidence in that object's degree of existence. We need this to allow an object to have zero credence in a model (so it is not there at all) and we are completely sure that it is not there. These are clearly different concepts and they have different interactions within models and in forming expectations.

The intuition is that if one model predicts that X will be there and another that X will not be there then the situation is different than one where one model predicts that X will be there and another with no prediction about X. The former poses a conflict, the latter not; the difference is more nuanced than that discussed in the section on expectation above. The separation of the concepts of credence and credulity marks another sharp break with the BPD framework. Neither concept I use *is* propositional belief nor is the propositional belief concept a

necessary component of cognition. I have not yet encountered any methodology that utilizes separate measures for 1) the strength of the element's presence and 2) the strength of one's justification for that figure, but the idea has great promise.[26] At present credulity is largely excluded from the construction of the model-based framework, but clearly its role in expectation, association, and imagination needs a more thorough exploration. This section serves to acknowledge that such a concept is relevant and included, but that space and time are limited.

5 APPLICATIONS OF THE FRAMEWORK

5.1 EXPERTISE AND HEURISTICS

Most people can juggle 5±2 ideas in their minds at once (reference is Kaplan, but I forget where).[27] Perhaps that is just the conscious limit and much more is going on "behind the scenes". Or perhaps we just switch which ideas we are considering very quickly. What is for certain is that people can develop expertise in specific subject matters to increase the speed and number of ideas that they can consider. One way to interpret expertise is having streamlined models that include only those elements that experience has deemed worthy of consideration. So it would not be the case that an expert has learned to juggle more ideas, but that an expert has learned which 5±2 ideas are sufficient to solve the problem, make the decision, or perform the task required. Furthermore, this seems to happen pre-perception; only the relevant ideas get processed into perceptions just like the way our visual sensory mechanism censors input before generating perceptions.

These simplified models (often called *heuristics*) provide shortcuts for cognition and vastly increase the range of skills that can be learned. Imagine having to consider every detail every time you wanted to ride a bicycle, solve a calculus problem, buy dish soap, or perform

[26] Expanding on this point takes us away from philosophy into (the philosophically relevant) fields of artificial intelligence, statistical reasoning, computer modeling. But for the sake of thoroughness I feel compelled to mention that Conceptual Slipnets, Bayes Nets, Semantic Webs, Dempster-Shafer Systems, Epistemic Networks, and many more situational modeling techniques (all of which use variations on belief) may benefit from a reworking and rethinking in terms of credence and credulity. Since **those** techniques are my area of expertise the insights gained from this philosophical exercise has provided me a plethora of new and exciting research projects.

[27] I don't mean to imply that people can only maintain 5±2 elements in their working models. That 5±2 notion is a common one in psychology and I find it rather humorously uninterpretable. I deploy that notion only to initiate the discussion of what limitation expertise is supposed to overcome and then describe how an analogous role is available in the model-based framework.

heart surgery. The first several attempts require a great deal of time and consideration of the options to learn how to do it, but eventually only one option for the next step even rises to the mind. This partly explains the aforementioned (§2.2.4) difficulty that experts have in describing why or how they do what they do. Of course everybody has expertise in many things, and the plausibility of this claim is bolstered once you recall (if you can recall) what it was like to *learn* to ride a bike or solve calculus problems or play an instrument, what it is like to just *do* those things after you've learned them well, and what it is like to try and *teach* those things to others.

This is not just meant to apply to procedural knowledge. We develop similar heuristics for how plausible certain occurrences, chains of events, combinations of objects, motions of objects, etc. are. There are nearly unlimited options for how one brings the credences of the myriad elements in one's models (and the credences of whole models) into congruence and cohesion in different contexts. Learning which elements "hang together" and how to assimilate new perceptions also admits to expertise. One may have well developed understanding of music theory as a stand-alone body of knowledge, but using that knowledge to interpret a performance or compose a concerto is something different. The propositional belief that (say) a G chord harmonizes well with the key of C does alone provide a desire to include G in the appropriate place in a sophisticated composition. In the model-based story at first one would imagine what Am, E, G, C, … chords would sound like in that part and then select the best sounding chord. Eventually one would have a vocabulary of relevant models for the C major key context and would simply have to select the models with the highest overall credence for this context. And only if that selection causes some cognitive dissonance would one have to evaluate more carefully the internal details of the model and/or the appropriateness of the model to the context and adjust the credence-context maps accordingly.

Thus what to perceive from our senses or imagination, how to associate perceptions to contexts, what objects, relations, dynamics, etc. get what credences, and how to adjust credences in light of new input are all aspects in which we gain expertise. And necessarily so – no creature could survive considering only a few concepts at a time and needing to consider everything every time. By encapsulating whole systems of concepts into modules we can handle ever more complex situations by implementing that whole module as an element of another model. We are binning the individual components into functional groups of a sort completely different from the boxologist.

5.2 MODULARITY

The modular nature of our models allows people to have drastically variant responses to perceptions in different contexts. One clear example is the laboratory scientist who is also a fervent churchgoer. Such a person has learned appropriate methods for analyzing causal structures, evaluating statistical data, and applying the scientific method, but does not apply them in every context. It is perfectly consistent with the model-based framework for cognition that a person could have two separate sets of responses each appropriate and well adapted to different contexts as long as those context rarely interact or interact very weakly. So that faithful scientist need not be tempted to use faith in research explanations or attempt to experimentally investigate religious doctrine; if the contexts are separate enough the faith module will never even be activated in the scientific context.[28]

Hence people make different jokes with their family, friends, and employers. People speak at different volumes in different environments. And people generally adapt to their environments automatically without conscious consideration about what is appropriate where and with whom. Such contextuality is very difficult to explain under the BDP framework. The difficulty rarely comes up in the relevant discussions because there is an implicit assumption that some kind of context regulating devise (also in terms of belief-desire propositions) is available and examples within BDP frameworks are very limited in scope. Credence is a local phenomenon; it only has to maintain coherence and consistency with connected concepts (i.e. within the module). Context dependence is entailed by the model-based framework because models are modular in the way described above and perceptions call models to the set of working models through learned associations (as described above). The associations between context-based perceptions (prompts) and the models they evoke may also help explain why some models generate behavior and others do not (discussed elsewhere).

[28] This is not meant to imply that the scientist ought to keep them separate or that he/she can maintain this separation without some cognitive cost or difficulty. I merely wish to explain a common occurrence of strongly incompatible mindsets being simultaneously held in terms of modular cognition. Evaluating such a practice is in the domain of morality and sociology.

5.3 MODELS AT WORK (AND PLAY)

Within the model-based framework creating fictional worlds happens very much in the same way that we fill in details of our real lives. In fact, representing the mind with models blurs many of the distinctions made by others working on explaining how fictional worlds are created. Here I will simply focus on the point that we need not make the (implausible) claim that to contemplate a fictional world people generate a whole duplicate real world representation and then change the necessary features to match the known elements of the fiction (Skolnik 2006, Walton 2006). Instead, the perceived elements of a fiction or reality seed the growth of a working model that brings in relevant and associated modules from the model vocabulary. This may happen in a *prompt cascade* where aspects of the immediate scene first generate (prompt) elements of a mental model, those elements have associations from previous experience and thus prompt further elements and modules from the model vocabulary, which then prompt more and more elements and modules, and so on. One corollary of the prompt cascade idea is that one need not bring in whole worlds based on every prompt; one builds a working model (which *includes* the world under consideration as a portion) from only those aspects that are related to the prompt through some associational link.

For example, it does not make sense to ask if, in the fictional world of Batman tales, the construction of the Large Hadron Collider is underway at CERN in Switzerland. It also does not make sense to ask whether there has been an earthquake within the past month in Turkey. It does not make sense to ask what the interbank loan rate is in Gotham City. The reason is that we don't expect these details to be filled in for that fictional world even though we *do* expect them to be filled in for the real world. We might expect there to be an interbank loan rank for banks in Gotham because we might assume, if asked, that they operate similarly to real banks in US cities. And if the interbank rate somehow figured into a plot then we wouldn't be surprised to hear a figure for the interbank loan rate within a Batman story. But if an account[29] begins with the Joker robbing a bank we would still not automatically populate that world with an interbank loan rate, let alone a precise figure for it, let alone a Large Hadron Collider. Those

[29] I use 'account' to refer to any second-hand description of events, scenarios, scenes, objects, etc. whether fictional or real. It could be a movie, news report, biography, symphony (if program music), campfire tale, scale model, comic book, anything that is a (physical or abstract) representation of something that has happened or is happening or will happen (or existed, exists, will exist) rather than directly observing the occurrence of the events, scenarios, scenes, etc. These are communicated through linguistic material (see above) and produce models of the world through our linguistic (including symbolic) models. This may be a tangent, but I just wanted to be clear with my use of possibly ambiguous terms.

details just don't matter in generating our Batman world/model and we don't expect questions about those details to even have answers.

5.3.1 Play

The same prompt cascade idea may also help explain how play behavior is generated. First I divide play into two categories: pretending and games. Games provide plenty of prompts in the form of rules, categorization, randomizers, pictures, landscapes, figurines, etc. These prompts generate a working model that includes the appropriate elements for the game at hand. These game-based elements then elicit other elements through association: we would expect to find associations to other games, places in the world, times in history, fictional stories, mathematical concepts, color relations, and many other things. What gets brought into the working model will depend on the game at hand (e.g. different things for different board games, card games, sports, and role-playing games) and the player's experience. Clearly the prompts provided by the game alone are insufficient to generate a working model rich enough to indicate the appropriate behavior in game-play. *Some* other elements will need to be brought in such as mathematical modules, physics modules, fairness modules, temporal modules, etc. to fill in the game world and allow the imagination to determine the appropriate behavior (e.g. game move)

Pretense is slightly more nuanced to explain because frequently the prompts are not extrinsically based; pretending is sometimes prompted by "dreamlike" elements. If one "out-of-the-blue" starts pretending to be an elephant the model-based representation story goes as follows. The person first puts an elephant in a working model and associates herself with that elephant. Elephants will have some associated properties, behaviors, and relations (depending on experience) and those will be brought into the working model (depending on credence and relevance). The self will also have some associations and those will be brought into the working models as well. Then certain elements associated with the elephant will somehow become associated with certain elements of the self. I say 'somehow' because I don't have any strong ideas there and it might be a noncognitive function, but certain model-based considerations (such as physical similarities, functional similarities, similarities in names, etc.) will form some, if not all, of the associations of self and the target of pretense. Finally, one will take observations from the real environment and perceive them as objects in the working model such as they cohere with the other elements already in the working model (i.e. elephant as self in this example). Thus poles are perceived as trees and chairs are perceived as buckets within the

working model because those items jibe well with there being an elephant there (again such perceptions would depend one's experience of elephants and related things).

Most people's (including children's) cognitive modeling abilities are great enough to consider both the play world and the real world simultaneously – to keep both working models running in parallel with associations connecting which items are which between them (trees are poles, poles are trees). The "pole ↔ tree" link will be an element in the holistic working cognitive model and it will have a credence level. But remember that *credence isn't belief* and so under this representation it is not even hinted that one must believe or imagine *that poles are trees*. There simply exists elements in the "real-world" portion of the working model and the "play world" portion of the working model that are associated with each other at the moment (as long as those models are working). A person may react to elements in one model that are not in the other – such as moving out of the way of a car (that is not an pretend object in the pretense) or moving away from a snake (that is in the pretense but is not associated with anything in reality) – without having to switch working models or shift between belief and (old-meaning) imagination.

5.3.2 Science

Scientific theorizing offers a particularly clear example of how imagination plays an integral role in thinking about the real-world. Scientific understanding is encoded as models in the model vocabulary. These models include what objects exist, how they interact, how they are related, which elements imply which other elements (associations), and what contexts prompt which elements. Equations, theorems, diagrams, etc. are linguistic prompts for perceptions of elements of models. Those elements are imaginary until an experiment provides sense-confirmation of those elements. Two scientists are likely to have (at least slightly) differing working models even given identical prompts because of differences in how, when, why, and where the models were learned and differences in associations generated from other aspects of their backgrounds. Similar scientific models generate similar expectations and responses given similar prompts just as similar make-believe play models do. People separate practical everyday scenarios from scientific scenarios just as they separate practical everyday scenarios from play scenarios. And the parallel, yet mostly separated, cognition of working models provides one sort of explanation of how this happens.

6 CONCLUSIONS

The model-based framework presented here offers an alternative method to the classical belief-desire propositional framework for providing mental explanations of behavior. To do so many familiar concepts related to cognition were discarded in favor of ones that represent cognition as akin to multiple interconnected and parallel simulations of scenarios. The new approach, though unfamiliar, perhaps unusual, and not fully fleshed out, better addresses some of the problems aesthetic philosophers have been struggling with under the old BDP paradigm. It is not that those problems were solved as they were posed, but rather that with the new representation those problematic aspects simply do not arise. Naturally some new problems arise and some new features need explaining, but if these are more productive areas of inquiry than some progress has been made in shifting the focus of the investigation.

This exposition has been necessarily broad and abstract because it is introducing an entire framework; specific applications would not be comprehensible without this foundational account. The next steps are 1) to take a specific problem from the literature and translate it into the model-based framework and 2) to expand and refine the foundational account. As a separate project from those two philosophical ones I can construct an autonomous software agent that implements this framework as its cognitive architecture. Such a test would help defend claims I made regarding this technique's closer proximity to cognitive processes and hence more apt for generating observed behavior. Other individual components (such as the separation of credence and credulity from belief and the measuring and comparison of expectation) may have more immediate application in other projects both philosophical and methodological. I hope you found the above presentation to be comprehensible, conceptually rich, and sufficiently original. And naturally I anticipate your comments on areas where you thought otherwise.

BIBLIOGRAPHY

Bramson, Aaron (2007) "Models of Science and the Role of Causation" Available upon request, eventually at www.bramson.net. [I think it's generally weak to cite one's own unpublished work, but it is a tangent to the purpose of this paper and only mentioned for locating the support for my claims on causation, supervenience, and levels of organization.]

Brink, David (1989) *Moral Realism and the Foundations of Ethics* (Cambridge: Cambridge University Press).

BonJour, Laurence (1985) *The Structure of Empirical Knowledge*. (Cambridge, MA: Harvard University Press).

Bruner J. S. and L. Postman. (1949) "On the Perception of Incongruity: a Paradigm." *Journal of Personality* 18, 1949: 206-223.

Crick, Francis and Christof Koch. (1998) "Consciousness and Neuroscience" in *Cerebral Cortex* 8, pp. 97-107.

Egan, Andy and Tyler Doggett (2007) "Wanting Things You Don't Want" Forthcoming in *Philosophers' Imprint*. Available from http://sitemaker.umich.edu/egana/papers

Kandel, Eric, James Schwartz, and Thomas Jessell (2000) *Principles of Neural Science*, 4th edition (New York: McGraw-Hill).

Kim, Jaegwon (1984) "Concepts of Supervenience" in *Philosophy and Phenomenological Research,* Vol. 45, No. 2. Dec. pp. 153-176.

Jeannerod, Marc (1997) *The Cognitive Neuroscience of Action* (Cambridge, MA: Blackwell).

Lewis, David (1986) "Causal Explanation" in *Philosophical Papers: Volume II* (Oxford: Oxford University Press).

Nichols, Shaun and Stephen Stich (2003) *Mindreading* (New York: Oxford University Press).

Osherson, Daniel (1995) "Probability Judgment" in *An Invitation to Cognitive Science: Thinking Vol.3* eds. Edward Smith & Daniel Osherson (Cambridge, MA: MIT Press) Chapter 2, pp. 35-75.

Prinz, Jesse (2004) *Gut Reactions: A Perceptual Theory of Emotion* (Oxford: Oxford University Press).

Schroeder, Timothy and Carl Matheson (2006) "Imagination and Emotion" in *The Architecture of the Imagination*. ed. Shaun Nichols (Oxford: Oxford University Press) pp 19-40.

Skolnick, Deena and Paul Bloom (2006) "The Intuitive Cosmology of Fictional Worlds" in *The Architecture of the Imagination*. ed. Shaun Nichols (Oxford: Oxford University Press) pp 73-86.

Walton, Kendall (2006) *Mimesis as Make-Believe* (Cambridge, MA: Harvard University Press).

Wegner, Daniel (2002) *The Illusion of Conscious Will* (Cambridge, MA: MIT Press).