An Embodied Approach to Representational Change in the Nine-Dot Problem

December 15, 2013

Tim Sparer

University of Illinois, Chicago

Abstract

Insight phenomena, such as the nine-dot problem, are well documented but less well understood. The mechanisms of insight are not nearly as well understood as those in conventional problem solving. While there is debate over the exact mechanism(s) of insight, it may be that many conventional experimental techniques (such as think aloud protocols) interfere with the insight process. To avoid these issues, this paper adopts techniques from embodied cognition research, specifically gaze-manipulation and eye tracking, to investigate a classic insight problem, the nine-dot problem. To guide this investigation, a hierarchical model for representational change is presented (in appendix G), connecting eye movements, perception and prior experience to nine-dot performance. In this model "insight" arises from a particular interaction of top-down (prior knowledge, information processing mechanisms) and bottom-up (sensation and perception) elements. From this model it is predicted that certain eye movements, congruent with those expected when viewing the problem's solution, will facilitate solutions, while eye movements that reinforce misleading initial perceptual elements will hinder performance. Importantly, the model accounts for the non-monotonic nature of insight.

An Embodied Approach to Representational Change in the Nine-Dot Problem

This experiment aims to show that insight in the nine-dot problem arises from the interaction between ‘bottom-up’ factors (sensation and perception) and top-down influences (strategy, prior knowledge and interpretation of task instructions) mediated by embodied factors (where eye movement operationalizes this construct). This experiment uses manipulations of attention, through eye movements, to investigate the connections between the nine-dot’s multiple sources of difficulty (Kershaw, 2002; Kershaw & Ohlsson, 2004), themselves hypothesized to arise from a combination of bottom up and top down factors.

To frame the rest of the introduction, the basic experimental design is briefly noted here. The basic experiment is simple: Either before or during participants’ attempt to solve the nine-dot problem, participants are asked to carry out a seemingly unrelated task: attending to (and saccading towards) the appearance of a simple stimulus on a computer screen. The order of appearance of the dots is, however, not random. Rather, it is designed (in some cases) to facilitate problem solving by mimicking the order of moves required to solve the problem, while in other cases the order of appearance of the dots may actually inhibit problem solving. Here, using saccades to ‘hijack’ cognitive processes is considered a manipulation of embodied cognition.

The remainder of this introduction comprises three topics: A brief review of research on the nine-dot problem, a brief review of embodiment in insightful problem solving, and notes on the use and interpretation of eye-movement in problem solving.

The nine-dot problem is a problem consisting of a 3 x 3 grid of dots, and the instructions “Connect all of the dots, without lifting your pen from the paper, using only four lines” (see appendix A). The nine-dot problem has long been a favorite for problem solving research (Burnham & Davis, 1969; Chein et al, 2010; Chronicle, Ormerod, & Macgregor, 2001; Kershaw, 2002; Kershaw & Ohlsson, 2004; Lung, & Dominowski, 1985; Macgregor, Ormerod, & Chronicle, 2001; Maier, 1930; Maier & Casselman, 1970; Ollinger, Jones, & Knoblich, 2013; Scheerer, 1963; Weissberg & Alba, 1981). Its solution (a sort of triangle with a slash through it or bow and arrow (see appendix A)), though obvious once seen, is far from apparent to most problem solvers, leading to remarkably low solution rates. Qualitatively, the 9-dot problem bears the hallmarks of insight problem solving: a sharp increase in feelings of warmth just prior to solution (Chein et al. 2010) and repeated frustrated attempts at solution leading to impasse, followed (in successful cases) by the rapid realization of the correct solution (Ollinger, Jones, & Knoblich, 2013).

From the standpoint of this paper, the nine-dot problem is an ideal “model organism” for understanding the contributions of multiple domains (from semantic understanding through perceptual organization and embodied cognition) to problem solving. The simple problem statement, an image of nine dots, and a single instruction with only three constraints (continuity of movement, number of lines for a solution, number of dots covered), covers significant psychological real estate, from low level perception (the organization of the dots) to high level cognition (semantic interpretation of the rules using prior knowledge, selection of heuristics and strategies). It is this combination of simplicity and scope that makes the problem so useful, though also difficult to experimentally dissect.

The simplicity of the problem statement, yet subjects' great difficulty in solving it (Kershaw, 2002; Kershaw & Ohlsson, 2004), reflects this dynamic between simplicity and scope. Theories ranging from the original Gestalt (Maier 1930; Scheerer, 1963; Maier & Casselman, 1970) to neurophysiological (Chi & Snyder, 2012), information processing (Macgregor et. al, 2001; Chein et al, 2010), and re-structuring/representational change (Ollinger et al. 2013), to "nothing special" (Weissberg & Alba, 1981), have attempted to explain the contradiction between the apparently simple problem statement and solvers' poor performance. These explanations tend to favor certain sources of difficulty over others (e.g. information processing, v. perceptual sources), though as Kershaw and Ohlsson (2004) noted, multiple factors contribute to problem difficulty. Understanding the nine-dot's sources of difficulty and their mutual influence on one another provides a window into the organization of cognitive processes.

This paper also argues that the sources of difficulty in the nine-dot problem reflect a fundamental hierarchical organization of cognition. It builds on a theory that cognition is hierarchically organized, where abstract levels of cognition are composed of more elementary cognitive functions, a fact reflected both in theories of problem representation (Knoblich et al. 1999; Ollinger et al. 2013), cognitive processing (Gregory, 1998), compositional grammar (Battaglia, Borzenstein, & Bod 2012) and even structure/function relationships in the brain (Friston & Keibel, 2009; Battaglia, Borzenstein, & Bod 2012). For the purposes of the 9-dot problem, and in agreement with the preceding authors, it is suggested that this hierarchy proceeds from concrete, perceptual elements to abstractions such as strategies and semantic interpretations of rules (in part) by way of working memory, attention, and, embodied inference. This paper not only aims to establish the role of embodied processes in the nine dot problem, but further proposes to use techniques from research on embodied cognition (gaze manipulation) to probe the hypothesized hierarchical organization of cognitive processes at work in the nine dot problem.

A more specific description of a hierarchical model for cognition is provided in appendix G. However, before returning to a description of research on the nine-dot problem, a key feature of the proposed hierarchical model should be noted. The proposed model suggests that "higher level" cognitive functions (planning, semantic understanding etc.) arise from particular interactions of lower level component functions (e.g. perception), often mediated by embodied processes. This process, however, is not purely feed-forward or bottom up: lower level elements do not strictly entail specific higher-level interpretations in a one-to-one mapping. Rather, as part of systemic feedback loops, high-level cognitive functions can (to some degree) attenuate or suppress their lower level inputs. Through this process of preferentially altering (weighting) their own inputs based on prior activity/experience high-level cognitive functions can be said to impose particular "interpretations" on low-level elements by biasing those elements' activity. A simple example of this is Gestalt filling in, in visual illusions. For example, in the Kanisza triangle, (appendix B), low level percepts (the dark circles, edges, and empty spaces) are organized into an illusory figure by higher level perceptual expectations related to prior experience with shapes and contrast. The specifics of the model are further explained inappendices G and F. For now, the important thing to consider is that the hierarchical model is neither strictly feed-forward (bottom-up) nor top-down, but instead specifies the way in which hierarchical cognitive processes interact. (An example of interacting hierarchical processes in the nine dot problem, based on Ollinger et al's (2013) results is given in appendix H).

The nine-dot problem, review of experiments

The following review of research on the nine-dot problem reflects the hypothesized hierarchy of perceptual and cognitive functions in the nine-dot problem. Sources of difficulty and previously hypothesized solution mechanisms are considered relative to their hypothesized locations in a hierarchy of cognitive functions, beginning with the most primitive and proceeding to the most abstract; from perceptual mechanisms, to information processing, rule interpretation and overall problem representation/representational change[1]. Embodied factors (such as eye movments) are hypothesized to reciprocally interact with, and mediate between both bottom-up (such as perception and attention), and top down factors (e.g. strategy choice) in cognition. This review also highlights possible points of interaction between top-down and bottom-up elements through embodied processes.

In a forthcoming paper paper these hierarchical levels will be related to functional/ anatomical hierarchy in the brain, and connected[1] to Knoblich et al.'s insight that, ""knowledge structures tend to undergo local, peripheral, or superficial changes before they undergo global, central, or fundamental changes," (Knoblich et. al, 1999).

Perceptual impediments to insight in the nine-dot first attracted attention from Gestalt researchers (Maier, 1930; Maier & Casselman, 1970; Scheerer, 1963). In the original line of thinking, the perception of the dots as a square (a simple closed figure with good gestalt), implies a constraint on allowable moves: Participants' incorrectly infer that they should not "draw outside the (implied) box". In light of other factors and additional evidence (Macgregor et al. 2001; Maier & Casselman, 1970), this particular interpretation of the constraints implied by perception is somewhat debatable, though alternative interpretations of the role of perception and additional evidence imply a more nuanced role for perceptual factors.

While drawing a box around the nine-dots and instructing participants to work outside it, as done by Casselman and Maier (1970) did not improve performance on the nine-dot problem, several converging lines of evidence point to the importance of additional perceptual factors. Kershaw (2002) found that facilitating training using patterns similar to the shape of the solution improved performance. Macgregor (2001), in analyzing 12-dot variants found that participants performed better on variants with superior figural integrity.

Chronicle (2001) manipulated the figure-ground contrast around and between dots by placing empty circles around or shading underneath the original nine dots so as to suggest the proper solution shape. This manipulation did not seem to improve performance, and may even have inhibited performance by implying a figure-ground relationship between the dots and hint as opposed to implying the nature of the solution shape. Chronicle's finding may point to important interactions among perceptual factors, such as the competition between multiple interpretations of the stimuli afforded by different levels of perceptual processing (this theme, and the role of top-down and bottom up influences is explored later in the comments on Ollinger e. al's (2013) work).

The interpretation of evidence like Chronicle's (2001) indicates an unresolved ambiguity in perceptual explanations of difficulty in the nine-dot problem: What specific perceptual factors are at play, and precisely which perceptual factors are most important? This question was not sufficiently answered by the Gestaltists (Kershaw, 2002), and remains ill-defined. In keeping with both the hierarchical model developed in this paper and empirical evidence (Lee & Mumford, 2003; Claessans & Wagemans, 2008) perceptual processes themselves are here suggested to be hierarchically nested. This line of reasoning follows both anatomical and physiological evidence, as well as classic gestalt understanding (Claessans & Wagemans, 2008; Friston et al. 2012; Lee & Mumford, 2003; Van de Cruys & Wagemans, 2011). Heuristically: basic environmental factors such as contrast and location contribute to the recognition of contours, while contours in turn are assembled into rough shapes, and the details of those shapes are elaborated by higher order processes. Each of these higher levels, "interpreting" incoming signals through past experience, effects expectations about what "ought to be seen" at the levels below. This process roughly corresponds to a hierarchical ordering of principles invoked by gestalt psychologists; from figure-ground to proximity, co-linearity, grouping, and figure-completion (Claessans & Wagemans, 2008; Elder & Goldberg, 2002; Fang, Kersten, & Murray, 2008; Van de Cruys & Wagemans, 2011). As noted in the discussion of the Kanisza triangle, top down processes can act to attenuate or suppress bottom up information, sometimes resulting in gestalt-type perceptual illusions. (This ability of top down “a priori” inferences to alter lower level input might also explain the persistence of misleading perceptual cues (e.g. the perceived box)).

Importantly, in the proposed model, the perceptual inferences created by subjects’ internal processes (“I have inferred that these dots are connected by an (implied) line”), guide future action (“therefore I should saccade in the direction of that line”). This links participant’s inner model of the task to their actions in the world, and thus links the (model driven) side of embodied cognition to perception.

Understanding perception as hierarchically organized also resolves the ambiguity in perceptual explanations of difficulty in the nine-dot problem by allowing specific perceptual elements in the problem to be mapped to different levels of the perceptual hierarchy. The hierarchical perspective prioritizes the sometimes-competing claims as to the influence of specific perceptual factors. Contrast and figure ground relations explain Chronicle’s (2001) findings, as well as findings from a pilot study of Kershaw’s (2002). Co-linearity, adjacency, and figure completion in-part also explain Macgregor’s 12-dot (and other) results.

From the embodied perspective, this hierarchical interpretation of perceptual factors in the nine-dot problem is useful in two ways. When paired with a model of motor movements (as in Friston, Adams, Perrinet & Breakspear (2012)), hierarchical perceptual processes can be said to guide motor movements and drive attention, where salience is (in part) related to the hierarchical ordering of perceptual factors (more specifically, the “fit” of stimuli to internally generated perceptual expectations (derived from a solver's internal model of the task) may drive attention, and therefore movement (see Feldman & Friston (2010) for more on this)). For experimental purposes, the hierarchical model of perception may be able to specify the effects of particular types of movements/embodied interventions at different hierarchical levels; for example fast movements, corresponding to lower order processing (Ballard & Rao , 1997), may preferentially effect "lower order" (simpler, more easily computed) perceptual elements (for example, the contributions of contrast to contour recognition during a rapid saccade).

Performance on the nine-dot problem is not however, governed only by perceptual factors (Kershaw & Ohlsson, 2004; Macgregor, et. al. 2001; Ollinger et al. 2013). Top-down factors, such as heuristic use and strategy selection (Macgregor et. al, 2001), prior knowledge (Kershaw & Ohlsson, 2004; Lung and Dominowski 1985; Weissberg & Alba, 1981), and working memory capacity (Chein et. al, 2010) all influence task performance. These factors, and their interaction with perception, are best-understood using Newell and Simon's (1972) information processing approach to problem solving (a brief digression, clarifying this approach, follows).

In the information processing approach to problem solving, problem solvers are said to maintain a problem representation, "in their heads". This representation consists of four parts: an initial (or current) problem state and the (perceived) possible states, a goal state, operators for moving between states, and constraints on allowable moves. Strategies are sequences of moves, intended to move a problem solver from the current state to the goal state. The space of all possible moves is referred to as a search space. Heuristics are rules of thumb that often inform strategy choice; given the information processing limits of the problem solver (e.g. capacity limits, such as working memory), and the (often large) size of problem spaces relative to the solver's capacity, heuristics are used to make search through the problem space tractable. Information regarding problem states, operators, constraints, and potentially effective strategies, may come from current perceptions, prior knowledge or the combination of the two. (Shortly, this paper will explore the role of embodied factors in mediating between current perceptions and prior knowledge in the interaction of "top down" and bottom up information processes). Importantly, from the standpoint of the nine-dot problem, information-processing factors, such as heuristics, have been shown to account for difficulties over and above those suggested by perceptual accounts (Kershaw & Ohlsson, 2004; Macgregor 2001). That is, while "bottom-up" perceptions influence difficulty in the problem, top-down information processing factors, in the form of heuristics and rule interpretation, independently contribute to problem difficulty.

The first information-processing factor considered is problem solvers' (mis)interpretation of rules and constraints. Lung and Dominowski (1985) hypothesized that subjects incorrectly understood the rules to suggest that moves not start or end on non-dot locations, while Macgregor (2001) and Kershaw and Ohlsson (2004) demonstrated a similar problem in subjects' understanding of non-dot turns. All three authors, by providing training, and sometimes, explicit instruction, found evidence for these sources of difficulty, and thus were able to help subjects solve the nine-dot problem. Properly interpreting the rules and the constraints that those rules do or do-not imply is important in the nine-dot problem especially, because problem solvers past experience (e.g. connect-the-dots puzzles) may lead them to misinterpret the rules (Kershaw, 2002; Weissberg & Alba, 1981). Without a basic understanding of the problem constraints expressed in the rules (allowable moves etc.) problem solvers cannot be expected to form effective strategies, much less effectively solve the problem.

Even with an effective rule-understanding, the search-space of the nine-dot remains problematic: If solvers constrain themselves to the perceptually implied "box" they ignore the correct solution. If, however, they work "outside the box" there is no clear boundary to the search space, making the problem (nearly) intractable. Even within the implied box, the space of possible moves is quite large and therefore of questionable tractability. The search space itself can be usefully modified to some degree by explicit hints and visual cues[2] (Weissberg & Alba, 1981), or perceptual and representational cues (Ollinger, Jones, & Knoblich, (2013)[3]), or solvers' themselves may adopt heuristics to more effectively guide their search.

Even within the implied box, the size of the search space demands that problem solvers adopt heuristics to reduce it (Macgregor et al. 2001). Since the search space of the problem is so large and the amount of look-ahead required for solution is high Macgregor (2001) suggests that subjects use two heuristics to judge the value of a particular move: initially, problem solvers' evaluate a move based on the number of dots it connects (more dots equals better move). Unfortunately, this heuristic actually inhibits the correct solution. Those subjects' who correctly solve the problem must adopt a second

---

[2] Weissberg and Alba (1981) told participants that the search space was exhausted, and provided nine-dot diagrams with the first, and first and second lines drawn on them.

[3] Ollinger et al (2013) placed arrows indicating the solution paths on the individual dots.

heuristic, usually after repeated failures with the first. The second heuristic assigns value to a move based on the "ratio of the number of remaining dots to the number of lines available to cancel them," (Macgregor, 2001).

The role of embodied processes may be less obvious in these higher order cognitive domains, but it is hypothesized here that embodied processes mediate higher order cognition in two ways, both of which serve to alter the problem space: interroceptive inference (similar to Friedman and Forster's (2000, 2002) term "cognitive tuning"), and changing perceptual affordances. Interroceptive inference is the notion that movements (or rather internal, propioreceptive (and related) signals) themselves provide information about the world, independent of any other sensory perception enabled by those movements (Seth 2013). Perceptual affordances are the "opportunities to perceive" provided by a particular orientation in space[4]. Both processes may interact with "top down" factors in the nine-dot problem. Guided eye-movements, as interroceptive inference, may subconsciously prime subjects as to particular search strategies. Further, by guiding subjects' eyes to un- or under-explored spaces in the physical-problem space, increasing perceptual affordances beyond what subjects' internal model of the problem would otherwise dictate, guided eye movements may also help to appropriately enlarge (and/or constrain) the problem space. Both factors (perceptual affordances, and interroceptive inference) may drive attention to key misconceptions in subjects' problem

[4] For example, turning one's head from the left to the right does not physically alter the space that one is perceiving, but alters the "content" of perceptions by re-orienting the perceiver.

representations (e.g. non-dot turns, starting and stopping points, implied "edges" to the nine-dot box).

Importantly, the embodied approach, when paired with the hierarchical model of cognition, offers a framework to develop well-specified, testable theories of problem solving as a dynamic process. Problem solver's problem representations create expectations about effective moves through the problem space; physical exploration of that space (eye movements etc.) either confirms or disconfirms those expectations, and problem solvers (presumably) alter the knowledge structures/problem representation to account for any new information encountered in their problem solving attempts[5]. The hierarchical framework specifies how knowledge structures should change, while the embodied approach provides a window for viewing (and sometimes driving) those changes.

Embodied problem solving is an iterative process; given current (bottom up) perceptual input, top down expectations drive the selection of the next best move, while each move offers a new set of perceptual affordances on which to condition top-down expectations. A simple enough process, with, however, an interesting twist: Ballard and Rao (1997) point out that any complex process, made of smaller component processes, must necessarily take (at least additively) longer than those component processes. This means that the business of updating more complex knowledge structures (and deriving

---

[5] Precisely how such knowledge structures are altered is not fully addressed in this paper. However, the hierarchical Bayesian framework of predictive coding, as applied to Gestalt principles, as suggested in Van de Cruys and Wagemans, J. (2011), perhaps combined with predictive coding interpretations of active inference as in Brown, Friston, & Bestmann (2011), seems promising.

action plans from them) must take longer than that for basic perception/action loops[6]. If more complex mental processes play out on different time-scales than the simpler processes that comprise their inputs, then experimental techniques may be able to distinguish among hierarchically organized processes by the timing and duration of interventions.[7] An embodied intervention, such as gaze direction, does precisely this by manipulating a whole primitive input-output loop, which over time, nonetheless effects changes in more abstract knowledge structures (such as strategy selection, or constraint relaxation) at higher levels in the cognitive hierarchy. Cleverly applied, embodied interventions may prove an excellent complement to theories of hierarchically organized cognition. With this in mind, this paper now turns to specific considerations for embodied interventions in problem solving research.

Methodological considerations, and evidence for Embodiment in problem solving

The primary conceit of embodied cognition, that cognition is grounded in the physical structures of the body and its environment, either through interroceptive inference (Seth 2013) or perceptual affordances provided by the body, has attracted a great deal of attention (for example: Barsalou, 1999, 2009). Not only is embodied cognition a biologically realistic work-around for many computational and philosophic issues in problem solving (see appendix F for more), it opens the door to a number of useful experimental techniques as well. Since embodied cognition situates even the most abstract representations in environment/body/brain dynamics, it allows the experimenter

---

[6] Such between-level processing and interactions may account for the “non-monotonic” nature of insight.

[7] Moss et al’s work (2011) on the timing of hints and impasse may relate to this.

to probe deep relationships between actions and subjects' 'internal' abstractions. Nowhere is this more evident than in problem solving. Since actions can be seen both as physical outcomes of internal processes and as drivers of those processes (by providing information to such processes, either by actions' influence on perceptual affordances or interoceptive inference (Seth, 2013)), embodied approaches to problem solving can use actions as both independent and dependent variables, a useful methodological tool.

A number of studies have analyzed embodied components of problem solving. As an independent variable, movement has been used to (subconsciously) prime subjects in a number of problems. Thomas and Lleras (2007), following on work of Grant and Spivey (2003), found that directing eye movements in a solution-congruent pattern on Duncker's radiation problem increased subjects' chances of successfully solving the problem, even though subjects did not report recognizing any relation between the eye movement task and the radiation problem (though a follow-up study (Thomas & Lleras 2009a) found similar results for non-embodied covert shifts of attention). Thomas and Lleras (2009b) also found that subjects' performance on Maier's two-string problem could be improved by having participants swing their arms in a manner suggesting throwing the two strings together (the "insightful" solution); again, subjects did not seem to be aware of the connection between the tasks. Further afield, Cook, Mitchell, and Goldin-Meadow (2008) found that particular types of gestures improved children's ability to learn and understand mathematical concepts. Freidman and Foerster (2000, 2002), in line with the "cognitive tuning" hypothesis of embodiment (a variant of Seth's interoceptive inference (2013)) have found that arm flexion (indicative of approach behavior, possibly queuing curious, “open” cognitive processes) improves performance on a several insight tasks, while arm extension (indicative of avoidance behavior, possibly queuing focused “narrow” cognitive processes) seems to improve analytical processing but not insightful problem solving. Even factors well outside of conscious awareness, such as cardiac timing, can influence the most elementary cognitive processes, such as memory (Garfinkel, Barnett, Minati, Dolan, Seth & Critchely, 2013). The methodology adopted in the present study most closely resembles Thomas and Lleras’ (2007) approach to the radiation problem, and is explained briefly below, before considering the role of embodied factors as dependent measures.

Embodied eye movement in this study follows the lead of Thomas and Lleras (2007) by using participants’ eye movements to implicitly trigger solution congruent or solution incongruent internal representations. By manipulating the particular shapes traced by participants’ eyes (and the length of time spent attending to them), the study aims to analyze the impact of different levels of cognitive processing on participants’ solutions, and by extension, shed light on the levels at which constraint relaxation and/or problem elaboration are taking place.

Embodied interventions are not the only use of embodiment in problem solving research. Embodied processes such as eye movements, or other motor outputs, can serve as reliable non-verbal dependent variables. This is especially important in the case of insight problem solving, where verbalization can interfere with solutions (Schooler, Ohlsson, & Brooks, 1993). Grant and Spivey (2003) demonstrated differing characteristic patterns of eye movement in successful vs. non-successful insight problem solution attempts, while Knoblich, Ohlsson and Raney (2001) used fixation duration over the course of problem solving to distinguish between different phases of successful insight-problem restructuring: improper problem representation resulting in impasse, and effective constraint relaxation (following impasse): Different locations and durations of fixation appear indicative of different problem phases. Mirman and colleagues (Mirman et al, 2012; Stephen & Mirman, 2010), have also suggested that the aggregated distribution of gaze steps is important, and may reflect the interactions of different components/processes in cognition; given the hierarchical model suggested in this paper (appendix G), more traditional eye movement measures, such as those used by Knoblich et. al (2001) and Grant and Spivey (2003), might be useful for indicating the current level of cognitive processing, while measures such as Mirman's may prove crucial to understanding the interactions between levels.

A final approach to embodiment seeks to understand embodiment and cognition as a self-organizing dynamic system. In such systems, embodied dynamics are part of interacting agent-environment loops. Actions are the ouput of cognitive processes, which are themselves grounded in past embodied dynamics. This approach is, as the name implies, less concerned with outputs per say, and more concerned with dynamic interactions. It suggests that the role of cognition is to (dynamically) stabilize the organism in its environment, and that cognition is the result of multiple interacting and self-organizing processes both within and outside the cognizer. (Mirman et al. 2012,Stephen & Dixon, 2009; Stephen & Mirman, 2010). The current experiment draws on this approach to understand embodied dynamics and shed light on the proposed hierarchical model.

A specific contention, regarding eye movements, indicating dynamic processes, made by Mirman et al (2010; see also Stephen & Mirman 2012), is that the distribution of embodied measures (such as gaze steps) can be understood to reflect interactions in underlying cognitive processes. That is, different types of cognitive processes have different characteristic distributions: Additive, component dominated cognitive processes have Gaussian distributions, Multiplicative (interacting) componential processes have lognormal distributions, and processes characterized by heavy feedback loops show Power-law (scale free) distributions (Mirman et al. 2010, Stephen & Mirman, 2012). Mirman et. al further suggest that actor/environment dynamics play a key rule in regulating the type of processing that occurs: highly constrained circumstances (either due to task constraints or the abilities of the actor), tend to lead toward more stable, component driven processes (Gaussian or Lognormal) while less constrained states tend toward power law processes. Thus, the distributions of embodied measures (such as gaze steps) can be used to reason about internal or external constraints on cognition.

Mirman's findings (2012, Stephen & Mirman, 2010) as well as Stephen & Dixon's (2009) on hand-movements in insight problem solving suggest that, during constraint relaxation, eye movements (gaze steps) should exhibit power-law like behavior. This provides empirical traction for, and is consistent with, predictions from the hierarchical Bayesian (predictive coding) model outlined in the appendix G: As errors accumulate at one model level, feedback from other levels becomes more important. As

Mirman (2012) notes, the statistical signatures of feedback processes are power-laws. Thus, Mirman's interaction-dominant argument, past findings of power laws in insight, and the predictions of the Bayesian model (presented in the appendix) are in accord. Since more stable dynamics are indicated by lognormal distributions, the distribution of gaze steps might be expected to appear log-normal during initial problem solving and/or impasse, and as a power law distribution during insight.

Thus, this study uses elements of embodied cognition in three ways. As an independent variable, to probe various levels of the hypothesized cognitive hierarchy, as a dependent variable to suggest the level of the cognitive processing hierarchy currently occupied by the subject (similar to Knoblich, Ohlsson, & Raney, 2001), and as a dependent variable indicating movement and coupling between hierarchical model levels (in line with Mirman et al. (2010). The nature of embodied processes in a rough model of hierarchically organized cognition in the nine dot problem is considered more thoroughly in appendix G.

A final thought concerns distinguishing embodiment in the form of directed gaze, from broader notions of attention. Thomas and Lleras (2009a) have found that manipulations similar to the gaze manipulations used in the radiation task, but directing attention, rather than eye movements, resulted in similar improvements in problem solving. The current study does not tackle that claim head-on, but rather approaches the separation of attention and embodiment at the level of computation. Rao and Ballard (1997) point out that for any set of nested hierarchical computations, more intensive ("high level") computations must necessarily take longer then low level computations. In their arguments, attention precedes embodied inference, where the level (and time scale) of "embodied" computation is considered to be the time required to process a perceptual input and "decide" on the relevant motor output (and perhaps, receive perceptual feedback as to the results of that motor process) a process they term "deictic" (pointer based) computation. This process, takes longer than its constituent processes of attention and orienting. The empirical time-scales for attention (50 msc) and embodied (.3 seconds) computations derived by Rao & Ballard, then, can be used to distinguish between levels of processing on the part of participants.

Further, it is hypothesized that the content at the various levels of computation, is also hierarchically organized. The shortest (attentional) time-scales use the most basic environmental stimuli to process basic perceptual factors (e.g. contrast, distance, and adjacency), while slightly higher levels use the inputs from the basic perceptual factors to form predictions about contours, and shapes. At longer time-scales (embodied (.3 seconds) and cognitive time-scale (2-3 seconds) (Rao & Ballard) lower level perceptual factors combine with working memory to influence the selection of possible moves. Thus, it may be possible to separate the effects of attention from those of embodied processes by changing the timing of the appearance of stimuli. For these purposes, the current experiment presents stimuli at both fast (attentional) and slow (embodied) rates.

Conclusion

The present experiment then, aims to integrate and analyze the effects of embodied problem solving to elucidate hypothesized hierarchical interactions amongst problem components in the nine dot problem. To do so, the experiment manipulates gaze direction along various solution congruent and incongruent paths, each designed to take advantage of particular levels of hierarchical processing. Further, since hierarchical processes can be separated by times-to-compute, the experiment uses differing durations of stimuli/stimulus sequences to distinguish between levels of processing (attention vs. embodiment).

Methods

Participants

Four hundred student (10 x 2 conditions, 20 subjects each), recruited from the UIC subject pool will be used for this study.

Materials

Stimuli:

The nine dot problem (see appendix A), with instruction will be presented via an eye-tracker connected computer. The stimuli for the experimental manipulations, in the form of solution congruent, incongruent, or partially congruent sequences of dots/digit-strings as they appear on screen, can be found in appendix E .

Equipment

An Eyelink eye-tracking device will be used to present the stimuli and record the data. The experiment will be run in E-prime.

Procedures

This experiment consists of problem solving periods interspersed with “breaks” during which participants receive one of the experimental manipulations. During the initial problem-solving period, participants are given two minutes to attempt to solve the nine-dot problem (time derived from Kershaw & Ohlsson, 2004). It is presumed that this time period is long enough for participants to reach impasse. After the initial attempt, participants will participate in an eye-tracking task (see below) during which time they will receive one of the (2 x 10) experimental manipulations. Each manipulation consists of a sequence of dots which are either congruent, partially congruent, or incongruent with the 9 dot solution.

After completing the tracking task (which lasts for 5 seconds (at the longest) participants return to the nine dot problem. Thereafter, participants' solutions attempts are interrupted with the tracking task at 1 minute intervals until either 10 minutes have (has) passed or they have reached solution (whichever comes first).

Tracking task (adapted from Thomas & Lleras (2007)):

On an otherwise blank screen, a string of 7 digits containing either the number "1" or "2" appears. Participants will be instructed to press either "1 "or "2" on the keyboard once they have seen the digit. The duration for which a given digit appears at a particular location onscreen is part of the experimental manipulation. Depending on condition, dots may appear for a duration of either 50 msec (Rao & Ballard's (1997) attentional level), or .3 seconds (Rao & Ballard's (1997) embodied level). To ensure that participants do not simply use the stimuli as location cues, the first digit of the sequence will appear at a random starting point onscreen, followed (in the appropriate order) by the rest of the stimulus sequence. The task lasts approximately 5 seconds, and happens once a minute.

Since each stimulus consists of approximately 5 digit strings (in the case of the most complex stimulus, the “extended congruent” stimulus (Appendix E, stimulus 3, this is eight digit strings). Thus, at the longest time scale (the embodied scale), a complete stimulus presentation will take approximately 1.5 seconds (.3 x 5), or at the longest 2.4 seconds (.3 x 8). Since the task lasts for five seconds, participants will see multiple presentations of each stimulus. To avoid “giving away the game” after each sequence is finished, a digit string will randomly appear on screen, before beginning the next sequence. The initial position of the digit beginning the next sequence will also be randomized.

Proposed Analyses

To directly test the effect of the manipulations, both solution rates and overall solution percentage (per condition) will be assessed using ANOVAs. Planned comparisons, in line with the predictions about the impact of the interventions (see predictions) will be carried out for these measures as well.

Several other analyses will be used to analyze the eye-tracking data. It is predicted that those who successfully solve the problem will spend a larger amount of time (prior to solution) fixating on white areas outside the nine-dot space. This will be tested by partitioning the space into “within (implied) box” and “outside (implied) box” regions, and calculating a “time spent fixating on white space” ratio (time spent fixating on outside of box white-space divided by time spent on inside of box space). Non-solvers and solvers will be compared on this measure using a between groups ANOVA.

Within groups, the solvers white-space-ratios for the period prior to insight will be compared to the white-space-ratio during impasse.

Gaze step-size distribution will be analyzed using Mirman et. al's (2012) metric of "relative lognormality". Relative lognormality assesses the degree to which a subjects' gaze steps are best described by either a lognormal distribution (indicating a multiplicative component-wise process) or power law distribution (indicating a feedback driven process). This measure will be used for two comparisons, a between groups (solvers v. non-solvers) and within groups (solvers prior to and/or during impasse, and solvers just prior to insight/solution). In both cases an ANOVA (between groups, within) will be used.

Predictions

Expected solution rates (greatest to least). (Numbers in parentheses correspond to numbering of the stimuli in Appendix E).

> (3) Solution Congruent (Fully Extended) >= (2) Solution Congruent (Partially Extended) >= (1) Solution Congruent (Simple) >= (6) Ambivalent (Correct pattern, implied large box) >= (5) Ambivalent (Correct pattern, implied small box) >= (7) Interdot distance power law -2 >= (9) Random interference >= (10) Control >= (8) Inter-dot distance with power law less than 3 >= (4) Solution Incongruent

Expected solution rates (least to greatest) are the inverse of the above ordering.

These predictions follow from the level of constraint relaxation, and/or search space restructuring that each stimulus affords (see main text for further discussion).

Additional predictions:

Solvers will spend a great number of fixations on white space just prior to achieving solution than during impasse. Also, solvers will spend a greater amount of time fixating on white-space prior to solution than non-solvers will at a similar time-point in their problem solving attempts.

It is also predicted that solvers will also have a lower relative lognormality in the gaze-step distribution just prior to achieving solution than during impasse, reflecting more power-law like “(internal) cognitive interactions” due to expected feedback driven interactions “in subjects’ heads”. Also, solvers will have lower relative lognormality prior to solution than non-solvers will at a similar time-point in their problem solving attempts.

References

Akrami, A., Russo, E., & Treves, A. (2012). Lateral thinking, from the Hopfield model to cortical dynamics. *Brain research*, *1434*, 4-16.

Ash, I., Cushen, P., Wiley, J. (2009). Obstacles in investigationg the role of restructuring in insightful problem solving. *Journal of Problem Solving 2*(2) 6-41

Ash, I., Wiley, J. (2006). The nature of restructuring in insight: An individual-differences approach. *Psychonomic Bulletin and Review, 13*(1), 66-73

Ballard, D. H., Hayhoe, M. M., Pook, P. K., & Rao, R. P. (1997). Deictic codes for the embodiment of cognition. *Behavioral and Brain Sciences*, *20*(04), 723-742.

Battaglia, F.P., Borensztajn, G, Bod, R. (2012). Structured cognition and neural systems from rats to language. *Neuroscience and Biobehavioral Reviews 36*, 1626–1639

Baron, S. G., & Osherson, D. (2011). Evidence for conceptual combination in the left anterior temporal lobe. *Neuroimage*, *55*(4), 1847-1852.

Barsalou L.W. (1999) Perceptual symbol systems. *Behavioral. Brain Sciences*. 22, 577–660

Barsalou, L. W. (2009). Simulation, situated conceptualization, and prediction. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1521), 1281-1289.

Beuls K, Steels L (2013) Agent-Based Models of Strategies for the Emergence and Evolution of Grammatical Agreement. PLoS ONE 8(3): e58960. doi:10.1371/journal.pone.0058960

Beal, M. J. (2003). Variational Algorithms for Approximate Bayesian Inference. PhD. Thesis, University College London.

Bickhard, M. H. (2000). Information and representation in autonomous agents. *Cognitive Systems Research*, *1*(2), 65-75.

Bickhard, M. H. (2001). Why children don't have to solve the frame problems: Cognitive representations are not encodings. *Developmental Review*, *21*(2), 224-262.

Bickhard, M. H., & Terveen, L. (1996). *Foundational issues in artificial intelligence and cognitive science: Impasse and solution* (Vol. 109). Elsevier.

Bennett, C.H. (1998). Logical depth and physical complexity

Brown, H., Friston, K., & Bestmann, S. (2011). Active inference, attention, and motor preparation. *Frontiers in psychology*, *2*.

Battaglia, F., Borzenstein, G., Bod, R. (2012). Structured cognition and neural systems: From rats to language. *Neuroscience and Biobehavioral Reviews 36*, 1626–1639

Burnham, C., & Davis, K. (1969). The nine-dot problem: beyond perceptual organization. *Psychonomic Science, 17*, 321–323.

Cabessa, J., & Siegelmann, H. T. (2011, July). Evolving recurrent neural networks are super-turing. In *Neural Networks (IJCNN), The 2011 International Joint Conference on* (pp. 3200-3206). IEEE.

i Cancho, R. F., Riordan, O., & Bollobás, B. (2005). The consequences of Zipf's law for syntax and symbolic reference. *Proceedings of the Royal Society B: Biological Sciences*, *272*(1562), 561-565.

i Cancho, R. F., & Solé, R. V. (2003). Least effort and the origins of scaling in human language. *Proceedings of the National Academy of Sciences*, *100*(3), 788-791.

Claessens, P. M., & Wagemans, J. (2008). A Bayesian framework for cue integration in multistable grouping: Proximity, collinearity, and orientation priors in zigzag lattices. *Journal of Vision*, *8*(7).

Creutzig, F., & Sprekeler, H. (2008). Predictive coding and the slowness principle: An information-theoretic approach. *Neural computation*, *20*(4), 1026-1041.

Chi, R., Snyder, A. (2012). Brain stimulation enables the solution of an inherently difficult problem. *Neuroscience Letters* 515 (2012) 121–124

Chi, M. T. H (1992). Conceptual change within and across ontological categories: Examples from learning and discovery in science. In R.N. Giere (Ed.), *Cognitive models of science* (pp. 129-186). Minneapolis: University of Minnesota Press.

Changizi, M. A., Hsieh, A., Nijhawan, R., Kanai, R., & Shimojo, S. (2008). Perceiving the present and a systematization of illusions. *Cognitive science*, *32*(3), 459-503.

Changizi, M. A. (2008). The trade-off between speed and complexity. *Behavioral and Brain Sciences*, *31*(02), 203-203.

Chein, J. M., Weisberg, R. W., Streeter, N. L., & Kwok, S. (2010). Working memory and insight in the nine-dot problem. *Memory & cognition*, *38*(7), 883-892.

Chronicle, E. P., Ormerod, T. C., & MacGregor, J. N. (2001). When insight just won't come: the failure of visual cues in the nine-dot problem. *The Quarterly Journal of Experimental Psychology, 54A*, 903–919.

Colby, C. L., & Goldberg, M. E. (1992). The updating of the representation of visual space in parietal cortex by intended eye movements. *Science*, *255*(5040), 90-92.

Conant, R. C., and Ashby, W. R. (1970). Every good regulator of a system must be a model of that system. *International Journal of Systems Science 1*, 89–97

Cook, S. W., Mitchell, Z., & Goldin-Meadow, S. (2008). Gesturing makes learning last. Cognition, 106, 1047-1058

De-Wit, L. H., Kubilius, J., Wagemans, J., & de Beeck, H. P. O. (2012). Bistable Gestalts reduce activity in the whole of V1, not just the retinotopically predicted parts. *Journal of vision*, *12*(11).

Elder, J., Goldberg, R. (2002). Ecological statistics of Gestalt laws for the perceptual organization of contours. *Journal of Vision* 2, 324-353

Fang, F., Kersten, D., & Murray, S. O. (2008). Perceptual grouping and inverse fMRI activity patterns in human visual cortex. *Journal of Vision*, *8*(7).

Fass, D., Feldman, J. (2002). Categorization under complexity: A unified MDL account of human learning of regular and irregular categories

Feynman, R.P. (1972) Statistical Mechanics. Benjamin

Friedman, R. S., & Förster, J. (2000). The effects of approach and avoidance motor actions on the elements of creative insight. *Journal of personality and social psychology*, *79*(4), 477.

Friedman, R. S., & Förster, J. (2002). The influence of approach and avoidance motor actions on creative cognition. *Journal of Experimental Social Psychology*, *38*(1), 41-55.

Friston, K. (2009). The free energy principle, a rough guide to the brain? *Trends in Cognitive Sciences, 13*(7) 293-301

Friston, K., Breakspear, M., & Deco, G. (2012). Perception and self-organized instability. *Frontiers in computational neuroscience*, *6*.

Friston, K. (2011). Embodied Inference: or "I think therefore I am, if I am what I think*"*. *The implications of embodied cognition*. Eds Wolfgang Schacher, Claudia Bergomi. Imprint Academic

Friston, K. (2012). A Free Energy Principle for Biological Systems. *Entropy*, *14*, 2100-2121

Friston, K, Adams R.A., Perrinet, L., Breakspear, M. (2012). Perceptions as hypotheses: saccades as experiments. *Frontiers in Psychology* 3:151

Friston, K., & Kiebel, S. (2009). Predictive coding under the free-energy principle. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1521), 1211-1221.

Friston, K., Mattout, J., & Kilner, J. (2011). **Action understanding and active inference**. *Biol Cybern, 104 (1-2)*, 137-160.

Friston, K. J., Daunizeau, J., Kilner, J., & Kiebel, S. J. (2010). Action and behavior: a free-energy formulation. *Biological cybernetics*, *102*(3), 227-260.

Friston, K., Schwartenbeck, P., FitzGerald, T., Moutoussis, M., Behrens, T., & Dolan, R. J. (2013). The anatomy of choice: active inference and agency. *Frontiers in human neuroscience*, *7*.

Feldman, H., & Friston, K. J. (2010). Attention, uncertainty, and free-energy. *Frontiers in Human Neuroscience*, *4*.

Garfinkel, S. N., Barrett, A. B., Minati, L., Dolan, R. J., Seth, A. K., & Critchley, H. D. (2013). What the heart forgets: Cardiac timing influences memory for words and is modulated by metacognition and interoceptive sensitivity. *Psychophysiology*, *50*(6), 505-512.

Grant, E. R., & Spivey, M. J. (2003). Eye movements and problem solving: Guiding attention guides thought. Psychological Science, 14, 462-466.

Gregory, R. L. (1980). Perceptions as hypotheses. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, *290*(1038), 181-197.

Gregory, R. L. (1997). Knowledge in perception and illusion. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *352*(1358), 1121-1127.

Gick, M., McGarry, S. (1992). Learning from mistakes; Inducing analagous solution failures to a source problem produces later success in analogical transfer. *Journal of Experimental Psychology; Learning, Memory, and Cognition, 18* (3), 623-639

Hattori, M., Sloman, S., Orita, R. (2013). Effects of subliminal hints on insight problem solving. *Psychonomic Bulletin, Review 20*, 790–797

Hohwy, J., Roepstorff, A., & Friston, K. (2008). Predictive coding explains binocular rivalry: an epistemological review. *Cognition*, *108*(3), 687-701.

Jaynes, E. T. (2003). *Probability theory: the logic of science*. Cambridge university press.

Huang, Y., & Rao, R. P. (2011). Predictive coding. *Wiley Interdisciplinary Reviews: Cognitive Science*, *2*(5), 580-593.

Kelley, P., & Whatson, T. (2013). Making long-term memories in minutes: a spaced learning pattern from memory research in education. *Frontiers in human neuroscience*, *7*.

Kershaw, T. C. (2002). *What makes an insight problem difficult? Facilitating performance on the nine-dot problem*. Unpublished master's thesis, University of Illinois Chicago, Chicago, Illinois, United States of America.

Kershaw, T.C., Ohllson, S.O. (2004). Multiple causes of difficulty in insight, the case of the nine-dot problem. *Journal of Experimental Psychology: Learning Memory and Cognition, 30(1),* 3-13.

Kiebel, S. J., Daunizeau, J., & Friston, K. J. (2008). A hierarchy of time-scales and the brain. *PLoS computational biology*, *4*(11), e1000209.

Kirby S, Hurford J. The emergence of linguistic structure: An overview of the iterated learning model. In: Cangelosi A, Parisi D, editors. *Simulating the evolution of language*. New York: Springer Verlag; 2002. pp. 121–148

Knoblich, G., Ohlsson, S., Haider, H., & Rhenius, D. (1999). Constraint relaxation and chunk decomposition in insight problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(6), 1534.

Knoblich, G., Ohlsson, S., & Raney, G. E. (2001). An eye movement study of insight problem solving. *Memory & Cognition*, *29*(7), 1000-1009.

Kurtz, K., Lowenstein, J. (2007). Converging on a new role for analogy in problem solving and retrieval: When two problems are better than one. *Memory & Cognition 35* (2), 334-341

Kwisthout, J. (2012). Relevancy in problem solving: a computational framework. *Journal of Problem Solving, 5*(1), 18-33

Lakatos, I. (1975). Falsification and the methodology of scientific research programmes. In *Can Theories be Refuted?* (pp. 205-259). Springer Netherlands.

Lee, T. S., Mumford, D. (2003). Hierarchical Bayesian inference in the visual cortex. Journal of the Optical Society of America A 20(7): 1434-1448.

Lerner, I., Bentin, S., & Shriki, O. (2012A). Excessive attractor instability accounts for semantic priming in schizophrenia. *PloS one*, *7*(7), e40663.

Lerner, I., Bentin, S., & Shriki, O. (2012B). Spreading activation in an attractor network with latching dynamics: Automatic semantic priming revisited. *Cognitive science*, *36*(8), 1339-1382.

Lerner, Y., Honey, C. J., Silbert, L. J., & Hasson, U. (2011). Topographic mapping of a hierarchy of temporal receptive windows using a narrated story. *The Journal of Neuroscience*, *31*(8), 2906-2915.

Lim, B. C., & Klein, K. J. (2006). Team mental models and team performance: a field study of the effects of team mental model similarity and accuracy. *Journal of Organizational Behavior*, *27*(4), 403-418.

Limanowski, J., & Blankenberg, F. (2013). Minimal self-models and the free-energy principle. *Frontiers in Human Neuroscience, 7*

Lung, C.-T., & Dominowski, R. L. (1985). Effects of strategy instructions and practice on nine-dot problem solving. Journal of Experimental Psychology. *Learning, Memory, and Cognition, 11,* 804–811.

Luo, J., Niki, K., & Knoblich, G. (2006). Perceptual contributions to problem solving:

Chunk decomposition of Chinese characters. *Brain research bulletin*, *70*(4), 430-443.

Macgregor, J., Ormerod, T., Chronicle., E. (2001). Information processing and insight: A process model of performance on the nine-dot and related problems. *Journal of Experimental Psychology: Learning, Memory, and Cognition 27(1),* 176- 201.

Maier, N. R. F. (1930). Reasoning in humans. I On direction. *Journal of Comparative Psychology*, 10, 115–143.

Maier, N. R. F, Casselman, G., G. (1970). Locating the difficulty in insight problems: Individual and sex differences. *Psychological Reports, 26*, 103-117

Mayo, J. P., & Sommer, M. A. (2008). Neuronal adaptation: Delay compensation at the level of single neurons?. *Behavioral and Brain Sciences*, *31*(02), 210-212.

Mirman, D., Irwin, J. R., & Stephen, D. G. (2012). Eye movement dynamics and cognitive self-organization in typical and atypical development. *Cognitive neurodynamics*, *6*(1), 61-73.

Munuera, J., Morel, P., Duhamel, J. R., & Deneve, S. (2009). Optimal sensorimotor control in eye movement sequences. *The Journal of Neuroscience*, *29*(10), 3026-3035.

Murray S. O., Kersten D., Olshausen B. A., Schrater P., Woods D. L.*(2002). Shape perception reduces activity in human primary visual cortex. Proceedings of the National Academy of Sciences of the United States of America, 99(23), 15164–15169*

Moss, J., Kotovsky, K., & Cagan, J. (2011). The effect of incidental hints when problems are suspended before, during, or after an impasse. *Journal of Experimental Psychology: Learning, Memory, and Cognition*

Newell, A., & Simon, H. A. (1972). *Human problem solving* (Vol. 14). Englewood Cliffs, NJ: Prentice-Hall.

Ollinger, M., Jones, G., Knoblich. G. (2013). The dynamics of search, impasse, and representational change provide a coherent explanation of difficulty in the nine-dot problem. *Psychological Research,* 1-10

Patel D, Fleming SM, Kilner JM (2012). Inferring subjective states through the observation of actions. *Proc Biol Sci, 279 (1748)*, 4853-4860

Pulvermüller, F., & Fadiga, L. (2010). Active perception: sensorimotor circuits as a cortical basis for language. *Nature Reviews Neuroscience*, *11*(5), 351-360.

Ramachandran, V. S. (1996). The evolutionary biology of self-deception, laughter, dreaming and depression: Some clues from anosognosia. *Medical Hypotheses*, *47*(5), 347-362.

Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature neuroscience*, *2*(1), 79-87.

Rokeach, M. (1970). *Beliefs, attitudes and values*: San Francisco. Jossey-Bass.

Rao, R. P., & Ballard, D. H. (2004). Probabilistic models of attention based on iconic representations and predictive coding. *Neurobiology of Attention*, 553-561.

Scheerer, M. (1963). Problem-Solving. *Scientific American*, 208, 118–128.

Schooler, J. W., Ohlsson, S., & Brooks, K. (1993). Thoughts beyond words: When language overshadows insight. *Journal of experimental psychology: General*, *122*(2), 166.

Schwartenbeck, P., FitzGerald, T., Dolan, R. J., & Friston, K. (2013). Exploration, novelty, surprise, and free energy minimization. *Frontiers in psychology*, *4*.

Seth, A. (2013) Interoceptive inference, emotion, and the embodied self. *Trends in Cognitive Science* (11), 565 – 573.

Shanahan, Murray. (2009) "The Frame Problem", *The Stanford Encyclopedia of Philosophy* (Winter 2009 Edition), Edward N. Zalta (ed.), accessed at http://plato.stanford.edu/archives/win2009/entries/frame-problem/

Siegelmann, H. T., & Fishman, S. (1998). Analog computation with dynamical systems. *Physica D: Nonlinear Phenomena*, *120*(1), 214-235.

Simon, H. (1978). Information processing theory of human problem solving. In W. K. Estes (Ed.) *Handbook of Learning and Cognitive Processes, Volume 5*, (pp. 271-287). Hillsdale, NJ: Erlbaum.

Smith K, Wonnacott E. Eliminating unpredictable variation through iterated learning. Cognition. 2010;116:444–449.

Song, S., Yao, H., Treves, A. (2014). A modular latching chain. *Cognitive Neurodynamics 8*, 37–46

Stephen, D., Dixon, J. (2009). The self-organization of insight. *Journal of Problem Solving, volume 2(1),* 72 -101

Stephen DG, Mirman D (2010) Interactions dominate the dynamics of visual cognition. Cognition 115:154–165

Thomas, L. E., & Lleras, A. (2007). Moving eyes and moving thought: On the spatial compatibility between eye movements and cognition. *Psychonomic bulletin & review*, *14*(4), 663-668.

Thomas, L. E., & Lleras, A. (2009a). Covert shifts of attention function as an implicit aid to insight. *Cognition*, *111*(2), 168-174.

Thomas, L. E., & Lleras, A. (2009b). Swinging into thought: Directed movement guides insight in problem solving. *Psychonomic bulletin & review*, *16*(4), 719-723.

Tomasello, M. (2008). *Origins of human communication*. Cambridge: MIT press.

Tetzlaff, C., Kolodziejski, C., Markelic, I., Worgotter, I. (2012). Time scales of memory, learning, and plasticity. *Biological cybernetics: Advances in computational neuroscience. 106*, 715-726

Treves, A. (2005). Frontal latching networks: a possible neural basis for infinite recursion. *Cognitive Neuropsychology*, *22*(3-4), 276-291.

Vosnaidou, S. (1994). Capturing and modeling the process of conceptual change. *Learning and Instruction, 4* 45-69.

Van de Cruys, S., & Wagemans, J. (2011). Gestalts as Predictions- Some Reflections and an Application to Art. *Gestalt Theory 33, No.3/4*, 325-344

Verhoeff, T. (2013). The origins of duality patterning in artificial whistle languages. *Language and Cogntion 4*(4), 357–380

Wareham, T. (2012). What can (and can't) make problem solving by insight possible? A complexity theoretic investigation. *Proceedings of the international conference on cognitive modeling.* Rubwinkel, N., Drewitz, U., van Rijn, H. (Eds.)

Weisberg, R. W., & Alba, J. W. (1981). An examination of the alleged role of" fixation" in the solution of several" insight" problems. *Journal of Experimental Psychology: General*, *110*(2), 169.

Wu, L., Knoblich, G., Wei, G., & Luo, J. (2009). How perceptual processes help to generate new meaning: An EEG study of chunk decomposition in Chinese characters. *Brain research*, *1296*, 104-112.

Wu, L., Knoblich, G., & Luo, J. (2012). The role of chunk tightness and chunk familiarity in problem solving: Evidence from ERPs and fMRI. *Human brain mapping*.

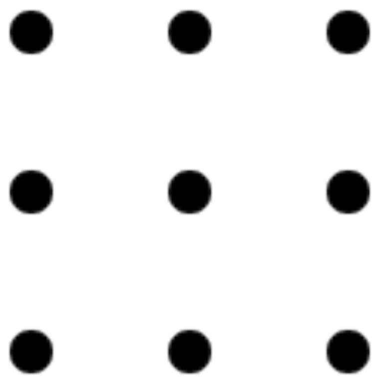
Wu, Q., Wu, L., & Luo, J. (2010). Effective connectivity of dorsal and ventral visual pathways in chunk decomposition. *Science China Life Sciences*, *53*(12), 1474-1482.

Zuidema W, de Boer B. (2009) The evolution of combinatorial phonology. *Journal of Phonetics, 37*(2), 125–144

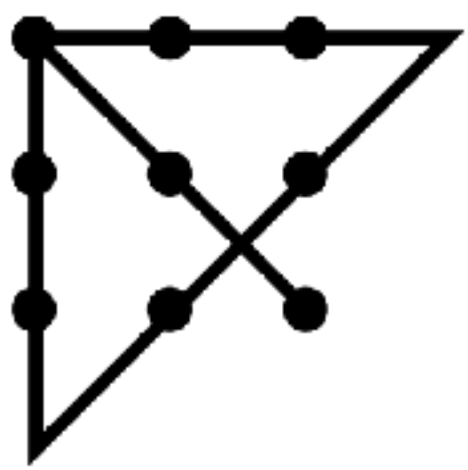
Zuidema, W., Westerman, G. (2003) Evolution of an Optimal Lexicon under Constraints from Embodiment. *Artificial Life 9*, 387–402

Appendix A: The nine dot problem and Solution

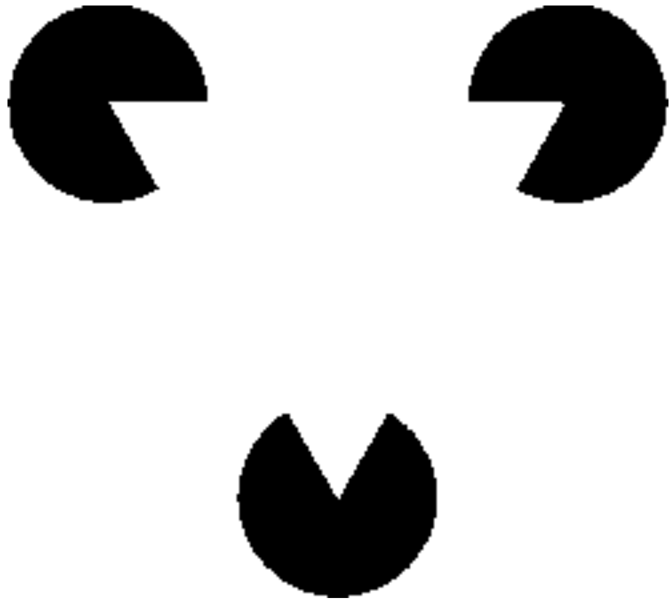
Instructions: connect all 9 dots (below) with four continuous straight lines without moving your pen from the paper.



Example Solution:



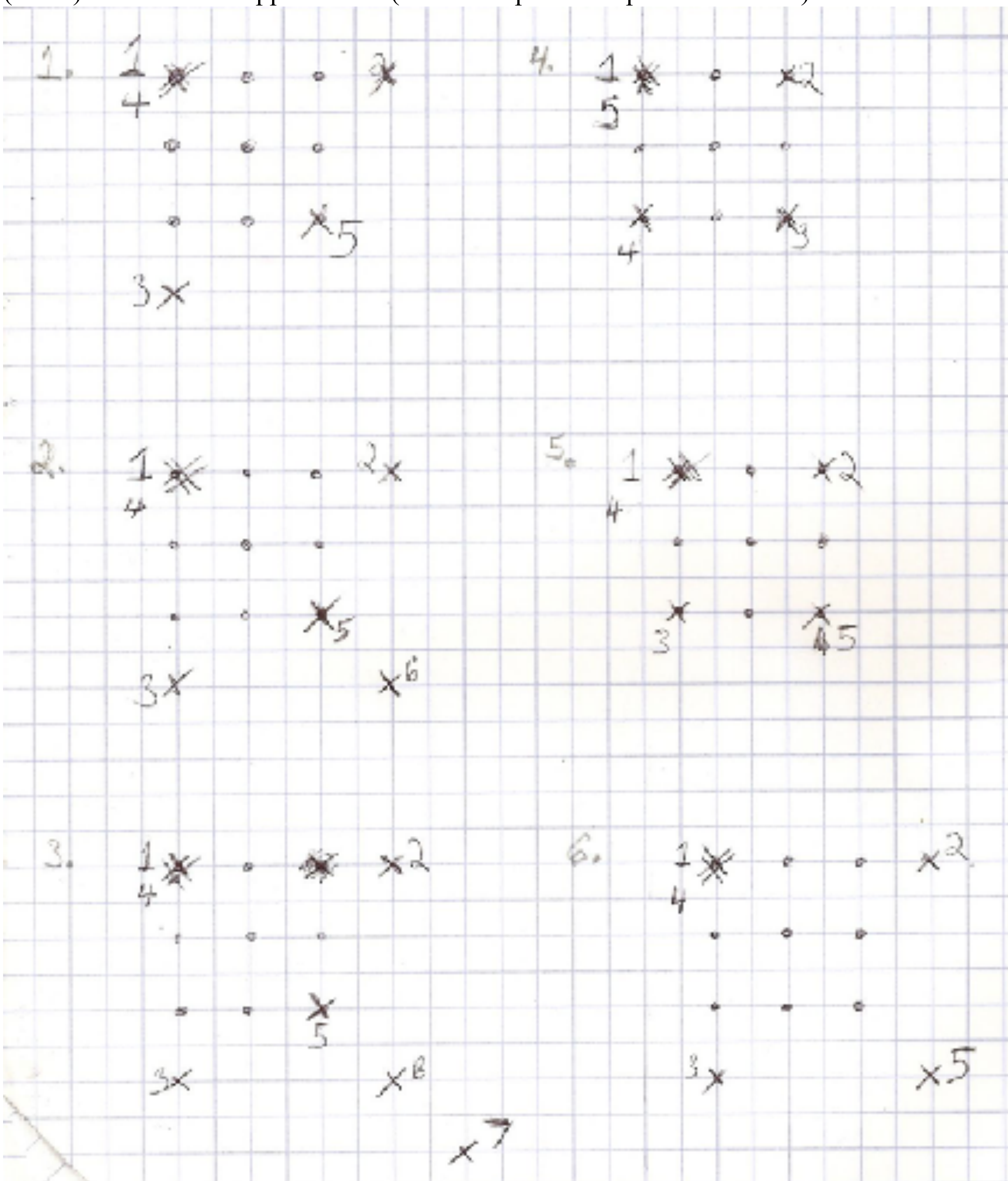
Appendix B: Kanisza Triangle



Appendix C: "My Wife or My Mother in Law" visual illusion

Appendix E: Example Stimuli Patterns:

Grids, and underlying '9 dots' are provided for reference, and do not actually appear in the stimuli. Only the numbered dots (marked with Xs) appear. Dots are numbered in the (serial) order of their appearance. (Further explanation provided below).



Stimulus sequences, further explained:

1. Solution Congruent (Simple): This pattern is exactly congruent with the solution.

2. Solution Congruent (Partially Extended): This is the same as pattern one, with the final (implied line) dot trajectory extended by one unit.
3. Solution Congruent (Fully Extended): This is the same as pattern one, but the final implied line/dot-trajectory is extend by two dots.
4. Solution incongruent: This pattern never breaks the 'square' implied by the dots.
5. Ambivalent: The correct pattern of eye movements is implied by the dot sequence, but the dots (and therefore, perhaps, implied search space) is constrained to the space of an implied box.
6. Ambivalent: This stimulus sequence presents the correct pattern, but extends the final dot sequence past the edge of the implied box (creating overall, an implied, "expanded box").

Additional sequences, not illustrated above.

7. Randomized inter-dot distance governed by power law, with an exponent of -2 . (inter-dot distance between current dot and next dot = dist^-2)
8. Randomized inter-dot distance governed by power law, with an exponent of --10 . (inter-dot distance between current dot and next dot = dist^-10)
9. Uniformly random interdot distance.
10. Control (blank screen).

Note: The power law exponents for number's 8 and 9 were selected to be lower and higher than the contour integration/co-linearity exponent of -2.92 suggested in Elder and Goldber 2002. These are to test the hypotheses about constraint relaxation at the level of colinearity and contour integration.

Appendix F: Theoretical Considerations for Embodiment as a driver of Restructuring

In the haste to cover all of the nine-dot's empirical bases, the discussion on representation and representational change through restructuring omitted several conceptual problems common to theories of representation and representational change. These include such knotty philosophical and computational issues as the frame problem and computational (in)tractability (Bickhard, 2000, 2001; Bickhard & Turveen 1996; Kwishtout, 2012; Wareham, 2012) . These concerns are not meant as idle philosophical pastimes; human problem solvers in the "real world" rarely seem troubled by such issues. It is this disconnect, between certain constructs in theories of representational change and human behavior that motivates the following discussion of the (conceptual) role of embodiment in representational change. This section makes a philosophical and computational case for mechanisms of embodiment in problem solving. The section thereafter presents some evidence for embodiment in insight problem solving (especially eye movements), and finally, the last section contains an integrated, hierarchical model for problem representation, combining both representational factors and embodied processes.

To begin with, some simple questions: Why should some problem features be more relevant than others, or some heuristics more obvious than others? How is relevance determined in the first place? What determines a basic problem "element" (as opposed to, for example, a percept)? Why should one problem representation (manner of organizing problem elements and determining relevance of individual elements) be chosen over another? This seems a bit similar to the discussion of gestalt perception, comparable to explaining why certain elements ‘pop-out’ or recede in a figure. Unlike the earlier discussion of Gestalt perception, at the level of representational change there are no “higher levels” of cognitive organization to appeal to. In this sense, representational change seems emergent[8] and self-organizing[9] (see for example, Stephen and Dixon (2009)). It is hypothesized here that the operations of representational change are emergent properties of the interaction of perception, prior-knowledge, and information processing activities with the environment, all carried out by the shared physical mechanisms of the embodied computational agent.

A detailed sketch of the embodied computational agent will be included in a separate theoretical paper on embodied problem solving and its mechanisms. Here three key points are provided: First, organisms (problem solving humans included) are open dynamic systems maintaining themselves far from equilibrium (Bickhard and Turveen, 1996)[10]. Second, by way of Conant and Ashby’s good regulator theorem and Karl Friston’s Bayesian free-energy principle, organisms, as self-maintaining dynamic systems, can be modeled as optimal Bayesian inference machines. Third, as embodied Bayesian learning, problem solving can be modeled as a hierarchical Bayesian inference process, beginning with perceptual activity, moving up through action and action

[8] Not entailed, but enabled, by simpler, lower-order processes.

[9] A process where global order arises out of local interactions in an otherwise disorganized system.

[10] Bickhard’s arguments, detailed below, are used to sidestep several troubling issues with classical problem solving accounts including the frame problem.

observation, and ultimately connecting to abstract semantic knowledge (this last point is considered in detail in the rough model of the nine-dot provided later).

Arguments for Embodied Representation

In order to maintain themselves, organisms must contain a model of themselves and their environment, as proved in the good regulator theorem of Conant and Ashby (1970) (also, Friston, 2013))[11]. However, Bickhard (2000, 2001) points out, that for basic philosophic reasons (including the frame problem and the problem of verification) such a model cannot be a purely correspondence or encoding based model, and instead must be based on the embodied dynamics of the organism. Rather than abstract representations, these dynamics map between sensory indications for action possibilities, and actions, where actions are selected based on their expected ability to maintain the organism. This "interactionast" account of representation sidesteps the frame problem, since the organism's model is not an explicit correspondence to states of some external world, but only a model for action based on sensory indication/differentiation. Rather than explicitly represent environmental states by internal correspondence, the organism implicitly differentiates environmental states based on the possibilities for action they afford. In this way the organism does not have to select among an infinite number of (possibly) true/relevant propositions representing environmental features[12]: Whatever relevant features the environment contains are implied in the pattern of sensory

---

[11] The good regulator theorem, a result from control theory, states that *every good regulator of a system must contain (or be) a model of the system regulated.* This holds for all homeostatic regulatory systems, including self-regulating ones. (Conant & Ashby, 1970).

[12] A version of the frame problem

indication/differentiation and the actions afforded by the indicated/differentiated environment[13]. Since actions are selected in order to maintain the organism, the interactionist approach connects the idea of environmental relevance[14] to self-maintenance through action selection. Importantly, this approach is biologically plausible, as recent work has shown that even basic ergodic dynamic systems with Markov blankets[15] can act as these sorts of self-maintainting systems (Friston, 2013).

The idea of embodied control has been formally mapped to Bayesian inference processes (Friston 2011; Friston, 2012; Friston 2013; Limanowski & Blankenberg, 2013; Seth 2013). This view casts Bickhard's self-maintaining systems in probabilistic terms. In this interpretation, each organism is represented as a specific probability distribution over a bounded subset of all possible environmental states[16]. In keeping with the good regulator theorem (described previously), organisms are hypothesized to contain (or embody) this same probabilistic model of (themselves[17] and) their environment[18]. As

[13] This also addresses the problem of verification: Correspondence/encoding accounts of representation, according to Bickhard (1996, 2000, 2001), cannot adequately address the problem of verification. In the interactionist account, a proposition about a putative environmental state is verified or falsified (only) insofar as the expected afforded action is successful (or not) in maintaining the organism.

[14] An important stumbling point in many accounts of problem solving, see Kwisthout (2012), and Wareham (2012).

[15] Such systems can be used as model's of the "chemical soup" at the origins of life. For a broader perspective on biological systems as dynamic systems see Friston (2012)

[16] The environmental state-space can be defined as all possible configurations of matter and energy in the the specific physical environment under consideration).

[17] Described as a probability distribution over environmental states.

[18] Arguably, the bulk of this "model" is contained in the dynamics of the organism's brain.

ergodic systems, the long run average of the time an organism spends in each state should equal the weight of that state in the organism’s probabilistic model. This means that organisms’ actions tend (or, are intended) to realize the distribution of states described in their probabilistic model. This formalizes, in probabilistic terms, Bickhard’s idea that organisms act to maintain the conditions necessary for their continued existence, and formally connects the organism as a physical structure to its dynamics (actions) over time. Put simply, organisms as self-maintaining systems ‘try’ to stay in the states considered most likely by their internal model.

The point of these mathematical moves is to bring the machinery of probabilistic inference to bear. The organisms’ internal model can be interpreted as a set of Bayesian priors, which specify the (prior) probability of the organism being in certain states of the world. Perception can be understood as inference about (hypothesized) external states given current sensations, with the organism’s internal model providing the priors needed for Bayesian inference. Actions are taken to minimize the deviance between the inferred environmental states and the organism’s desired states as specified by the prior probabilities of those states in the organism’s internal model, (where the model also specifies a “state -> action ->state” mapping)). Mathematically, this corresponds to minimizing the “surprisingness” of (the) environmental states that an organism finds itself in, relative to the organism’s probabilistic model, where surprise corresponds to the negative log probability of a particular state: Less probable states are more surprising. More simply, this says that in order to remain alive, organisms should avoid states that are improbable relative to their internal models: Fish stay in water, polar bears in the arctic. [19]

As Bickhard pointed out, organisms do not have direct access to hidden states of their environment, and must instead infer them through changes in sensory indicators. This implies that organisms cannot directly know the "surprisingness" of environmental states, and must instead infer it from the behavior of sense indicators. Therefore, organisms do not measure surprise per se, but rather assess an upper bound on the time average of surprise, known as "free energy" (Beal, 2003; Friston, 2011; Friston 2012; Friston 2013). The free energy represents how much the current (inferred) environmental state deviates from the expectations of the organism's model. By minimizing free energy, organisms ensure that they stay within the acceptable states prescribed by their internal model (the states necessary for their continued existence)[20]. An organism can minimize its free energy either by altering its model (perception/perceptual learning as inference), or altering it's environment through action ("active inference"). Under simplifying assumptions, the free energy reduces to the prediction error (the difference between the

---

[19] The term free energy comes, in part, from Richard Feynman's work (Feynman, 1972): it can also be understood as a quantity that bounds the evidence for a particular model. It can be thought of as analogous to the Gibbs or Helmhotz free energy in statistical mechanics where the thermodynamic quantities of entropy and (weighted)enthalpy, are replaced by model complexity (measured in bits), and precision-weighted accuracy (goodness of fit) of the model.

[20] A simple formulation of "free energy" can be found in the Kullback Liebler divergence (KL divergence) between two distributions. The KL divergence assesses the difference between two distributions. Minimizing the KL divergence between a model and an environment minimizes the free energy between that model and environment (see Appendix A for a description).

model's predictions and the observed data), in which case, reducing free energy amounts to reducing prediction errors through action or learning.

The free energy principle has been used to explain a number of biological and neuroscientific findings (Friston, 2009) and even bounded rational decision theory and exploration/exploitation trade-offs (Friston et. al, 2013; Schwartenbeck et al. 2013). The free energy principle may also provide a unifying account for problem solving. As well as serving as an explanatory principle for different approaches to problem solving (embodied v. symbolic representation), the free energy principle can serve as an over-arching principle to connect the embodied and information processing/representational approaches to problem solving in a computationally and biologically meaningful way.

It has already been shown that even simple dynamic systems act to minimize free energy (Friston, 2013). Further, it is know that the basic computational machinery of the brain, attractor networks and Hebbian learning, also minimize free energy (Friston 2009). Insofar as the embodied approach to problem solving invokes dynamic systems and relies on interpretations of brain dynamics, it easily fits a free energy framework. On the information processing side, the Markov models used in the free energy and optimal decision literature (Friston et. al, 2013; Schwartenbeck et al. 2013), may be good models for strategy choice (where strategies are sequences of moves a along the Markov chain). The fact that these models implicitly address exploration versus exploitation makes them a good candidate for theories of strategy change as well (see also Song, Yao & Treves (2014) for models of strategy change in attractor networks).

Further, the free energy principle formalizes several arguments made as to the nature of the "language of thought", the abstract grammar/symbolic representation systems (implicitly) invoked by information processing explanations of cognition. Supposing such systems develop to maximize the predictive generalizability of a finite, recombinant set of symbols, such systems must trade-off the cost of representational accuracy against the generalizability of the chosen representation. This complexity accuracy trade-off is inherent in the free energy minimization. Notably, such trade-offs may take place within a grammar or representational system (as in the case of theory change, and perhaps category learning (Fass & Feldman, 2002)), or on the grammar/representational system itself over longer time scales (e.g. the evolution of languages (Beuls & Steels, 2013; i Cancho, Riordan, & Bollobás, 2005; I Cancho & Sole 2003; Zuidema & Westerman, 2003). For the purposes of representation in problem solving, the idea that problem solvers try to find solutions that can be stated simply (in the particular formalism used by the problem solver) and deviate minimally from past knowledge, while accurately achieving the problem solver's goals, captures intuitive notions of how problem solving proceeds (these criteria might also be used to define intrinsically "good" solutions to problems with ill-defined goal states or large, under-determined solution spaces). That this process might be captured mathematically as search/free energy minimization on some sort of free energy landscape (defined by one's prior knowledge and the information given by the problem) is promising.

It seems the free energy principle offers a good explanation of top down processes (bounded decisions, symbolic representation), and bottom up phenomena (the computational machinery of the brain), but how does it connect these into a theory of embodied symbolic representation? Ideally, any theory of embodied representation needs to bridge between the biological hardware of the brain, the actions of the embodied agent, and the effective abstractions that constitute the language of information processing. Can the free energy principle do so? Here, it is argued that symbolic representation emerges from quasi-stable structures (attractors) governed by non-trivial dynamics in the brain (e.g. Treves, 2005), and stable external signal->referent mappings shared by (embodied) computational agents, which together are jointly necessary and sufficient for any meaningful theory of symbolic representation (e.g. : Cancho, Riordan, & Bollobás, 2005; I Cancho & Sole 2003; Zuidema & Westerman, 2003). It is further argued that both (complementary) processes can be understood in terms of free energy minimization.

The case for a free energy principle for symbolic representation rests on two arguments: The first is: symbolic representation is an emergent[21] property of evolving neural (attractor) networks implementing the free energy principle through Hebbian learning (such as those found in the brain). The second: symbolic systems themselves minimize a free energy function between communicating agents, different from (but related to) the function minimized by the attractor network supporting symbolic computation within agents. The first argument is straightforward. Biologically plausible models of symbolic language as instantiated in the brain rely on the machinery of the brain; those models which rely on methods compatible with the free energy principle

[21] But not entailed: The argument for this relates to the social origins of language, and is not treated in this paper. Though it will be in another.

(essentially any attractor network or Hebbian learning scheme, as well as others[22]), naturally implement the free energy principle (Friston, 2009). Treves (2005), and Battaglia et al. (2012) present the most biologically convincing models for brain-based implementations of both infinite recursion (Akrami, Russo, & Treves, 2012; Russo & Treves, 2013; Treves, 2005), and semantic construction/parsing (Battaglia, Borensztajn, & Bod, 2013; see also Lerner, Bentin, & Shriki, 2012A & B), both key elements in any symbolic language/information processing scheme. The mathematical methods employed by both Treves et. al, and Battaglia and colleagues are compatible with free energy minimizing interpretations.

The work of Treves and collaborators, as well as Battaglia et al. suggest that representation and information processing dynamics are best understood as features in an emergent hierarchy of cognition, (the dynamics of which emerge from and are governed by the "free energy principle"). At the top of both abstract-cognition and anatomical/ physiological hierarchies, the non-trivial dynamics of cortical networks support the basic machinery for computational information processing and representation, recursion (Treves, 2005) and semantic memory (Battaglia et. al 2012). Specific semantic primitives may come from representations of the physical body of both selves and others (Barsalou 1999; Friston, Mattout, & Kilner, 2011; Limanowski & Blankenburg, 2013; Pulvermüller, & Fadiga, 2010; Seth 2013). This begins to suggest answer to the question raised at the beginning of this section, "How does a problem representation hang together?" A problem representation hangs together in the dynamics of the cortical

[22] E.g. Battaglia et al's expectation maximization in stochastic grammar (2012).

network: if each stable attractor state in a cortical network constitutes an element of the problem representation (e.g. Song, Yao, Treves 2014), and at least some local attractors map to somatsensory/movement primitives (Barsalou, 1999; Friston, Mattout, & Kilner, 2011; Pulvermüller, & Fadiga, 2010), and transitions between states are non-trivial, then the attractors and the mapping between them might constitute a "problem representation". The relative activity of different attractors, and their place in the network may (in part) determine the relevance of a given construct at a given time. Changes in activation across the network constitute classical information processing operations/representations ("current state", "goal state", "move", "constraints"). (similar to Song, Yao, and Treves' (2014), approach to strategy change). Though the dynamics of the Treves' so-called "latching" networks are non-hierarchical and recursive, they sit at the top of an information processing hierarchy, with inputs from hippocampal-episodic memory and other brain systems (where episodic memory itself sits atop a hierarchy of perceptual inputs, see Battaglia et al. (2012) for details[23]).

The second argument for embodied abstract representation as free energy minimization, is slightly more nuanced. This argument states that symbolic representation is not solely the product of processes internal to single individuals in isolation, but rather that symbolic representation (in part) emerges from interactions between individuals interacting in a shared environment. This idea has a strong pedigree

[23] The cortical network does not operate in isolation; inputs into it (especially) from the hippocampus/episodic memory, or in the form of other sensory motor feedback can alter it's structure (over time); this might provide a base for representational change externally driven representational change. Internally driven change may arise from Treves' latching dynamics (see Song, Yao, & Treves, 2014).

in theories of the evolution of language as a social tool (Tomasello, 2008), as well as in arguments from language games, and language evolution through iterated learning models (Kirby & Hurtford, 2002; Verhoef, 2013). In these cases, symbolic systems emerge from the shared process of understanding intentions (Tomassello, 2008), and/or the shared struggle to map signals to referents in an easily reproducible manner as in language games (Beuls & Steels, 2013; Der Vylder & Tuyls, 2006; Zuidema & De Boer, 2009) or iterated learning models(Kirby & Hurtford, 2002; Verhoef, 2013). The free energy argument enters the picture in several ways.

Recall that, rather than prediction error, or deviance from a model, free energy can (also) be stated in terms of Bayesian model comparison. Given a collection of observations to explain, a model's free energy is equal to the complexity of the model minus the model's accuracy. Free energy minimization formalizes Occam's razor: simpler, more accurate models are preferred. In terms of shared symbolic systems, this sort of free energy minimization leads agents to seek the simplest, most accurate mapping of signals to referents with which to model their shared world and (possibly) joint intentions. In order to coordinate action, speakers wish to maximize the accuracy of their statements and minimize the chance for confusion (complexity), while listeners wish to minimize the processing costs needed to understand a speaker's intent. The resulting symbolic communication systems minimize the free energetic costs by seeking the simplest most accurate shared model with which to represent agents' shared world(s). I Cancho and collaborators (2003; 2005) have found evidence for this in least effort models of language generation, while Lim & Klein (2006) have demonstrated in "real world" settings (interacting teams) that groups that have the simplest (most similar) and most accurate shared mental models (of their task and team dynamics) are most successful.

When additional physical parameters (for example embodied (physical) factors of communication) and cognitive constraints on the transmission of symbols, as in iterated learning models, are factored in, it can be argued that the resulting symbolic communication system minimizes a free energy function across all communicators (optimizes the ease of communication for both speakers and hearers) which is grounded in the peculiar cognitive and physical constraints shared across agents (an important point, to be considered in a separate paper, when defining the construct of "relevance"). Verhoef (2013) has found evidence for precisely this in experimental tests of iterated learning (see also; Smith & Wonnacott, 2010). (While the other models mentioned in the preceding paragraph do not make the free energy connection explicit, such work can nonetheless support a free energy interpretation of symbolic representation, as the mathematical functions used can be understood as objective functions minimized by Bayesian 'free energy minimizing' agents).

This second interpretation of shared symbol systems offers another route for understanding embodiment. Insofar as embodied factors constrain symbolic expression (Zuidema & Westerman, 2003), but also provide the shared, underlying machinery for communication (e.g. mirror neurons, which have also been explained in terms of free energy minimization, see Friston, Mattout, & Kilner (2011), and also Patel, Fleming, Kilner (2012)), and also serve as initial symbolic primitives (e.g. through movement or gesture), embodied factors sit at (or near) the base of shared symbolic communication systems, as well. (There is far more to say on this, but for now, this paper does not pursue this line of reasoning further).

This emergent communication system can form the basis of the abstract languages, instantiated in the brain dynamics, discussed earlier. When agents (human problem solvers) represent problems (to themselves) in terms of this language, problem solving takes the form of (free energy minimizing) search through the representational landscape afforded by the language. The language acts as an internal generative model, which specifies the probability of possible solutions. For example, in a remote associates type task, (it could be argued that) problem solvers search an associative probability landscape determined by the (learned) co-occurrence statistics of natural language. More abstractly, problem solving may have the goal of representing new information (or transforming external information) in the simplest manner possible[24] while accurately representing external circumstances, to achieve some desired outcome.

It appears that the combination of internal factors supporting the emergence of recursion and semantics (while minimizing internal free energy) and external factors, such as the need for social coordination (minimizing the complexity and maximizing the interpretive accuracy of a shared world-model/ communicative system), mediated by shared embodied factors, gives rise to the sort of internal representations thought necessary for problem solving. Although it may give rise to a form of symbolic system similar to those in classical information processing accounts, this embodied form of

---

[24] Simple relative to the expressive power of the problem solver's existing language(past knowledge represented in the problem solvers' existing lexicon).

representation has additional computational advantages, otherwise lacking in pure information processing models: It address the frame problem (Bickhard & Turveen, 1996). It has also been shown by Siegelmann and collaborators (Cabessa & Siegelmann, 2011; Siegelmann & Fischman 1998) that the types of computation carried out by evolving neural networks and attractor networks (such as those used in the network models described above) actually exceed the bounds of normal Turing machines, and thus, may very likely overcome the computational hurdles to representation pointed out by Kwisthout (2012), and Wareham (2012).

The important difference between strict information processing, and the representational system described in this section is the source of the representational system. Only by grounding representation in embodied dynamics is it possible to avoid problematic computational issues, and only by understanding these dynamics through Bayesian free-energy minimization is it possible to construct an idea of “optimal” representational systems and specify rules for (optimally) moving through a representational space. Thus, the free energy account can be used to show how embodied (brain/body) dynamics influence the shape of abstract representations (by constraining the development of the representational system), and how abstract representations (as instantiated in brain dynamics) influence behavior (by constraining actions over time (e.g. action planning), and by constraining mappings between internal representational states). This idea will be revisited in concrete form in the section on eye movement in the nine dot problem.

The broader point of this section was to argue that embodied/dynamic and classical information processing approaches to problem solving are not at odds with one another, and are in fact complementary. Both are components of the process of free energy minimization in embodied, socially interacting computational agents. Embodiment/dynamic approaches address the computational problems of pure information processing accounts, while information processing accounts (framed in terms of minimizing free-energy within a representational system) capture the "symbol-system-bound" nature of higher order processes such as goal setting and the determination of 'good' solutions to problems[25]. A more specific treatment of the free energy principle as applied to problem solving, will be provided later as part of a hierarchical Bayesian account of problem solving as predictive coding (a specific, hierarchical form of free energy minimization). For now, having established a theoretical connection between embodied problem solving and information processing/representational accounts, this paper turns to evidence for embodied factors at work in problem solving.

Appendix G: Putting it all together- embodied Inference in the nine-dot problem; a rough model, application and concerns for experimental design

A theory of embodied problem solving must connect embodied behaviors to high-level cognitive processes. The appendix (F) on embodiment and representation suggests a high-level, principled way to do so. The remainder of this section suggests a practical

[25] Where the goodness of a solution is a judged by it's complexity (simple is better), and accuracy (fit of the proposed solution within the constraints of the problem).

application of these principles in the context of the nine-dot problem, (a more explicit model connecting cognition and embodied behaviors in the nine-dot problem will be presented in an additional technical reference, currently not included here).

Classic accounts of problem solving (Simon, 1978) describe problem solver's internal understanding of a problem as a problem representation composed of an understanding of the problem's initial state, goal state, rules (allowable moves), and operators. This representation is composed of perceptual information about the problem combined with prior knowledge. Since experimenters do not have direct access to subjects' internal mental states, problem solvers' problem representations must be inferred from behavior. Mapping between problem solvers' hypothesized internal states and observed behavior requires a model of problem solvers' internal processes on the part of the experimenter.

In the case of embodied problem solving an experimenter's model must not only specify high level behaviors, (such as the expected sequences of moves in the nine-dot problem (as done by Macgregor et al (2001))) but must also account for low level motor behaviors, such as eye movements. When necessary and where possible, models of embodied cognition ought to explain the role of movement in shaping internal representations as well as the effects of internal representation (as prior knowledge) on movement.

This dual role of basic motor movements, at once shaped by and shaping subjects' internal models, can be explained by way of analogy to higher levels of cognition, namely hypothesis testing and theory revision. That is, movements reflect tests of subjects' internal model of the world, and that model is revised based on the sensory consequences of movement. In this understanding, perception is the process of inferring the causes of sensory stimuli based on past experience, and movements validate (or falsify) those inferences. Continuing the analogy with hypothesis testing, past experience provides the theory, perceptions are hypotheses, and movements are experiments (Friston et al 2012; Gregory, 1980; Gregory, 1998).

This analogy suggests a unifying construct for embodied problem solving: embodied problem solving as an inference process[26] over multiple temporal and spatial scales. At short time scales (e.g. the time scale of saccades) and with high spatial resolution (the extent of visual field) basic environmental elements are combined with prior knowledge to form hypotheses about the environment (e.g. "this collection of dots suggests a contour"[27]) which are tested by individual movements ("saccade along the edge of an expected contour to see if the contour continues, as predicted from past experience with contours")[28]. At longer time scales, perceptual elements may be 'chunked' together (based perhaps on physical or temporal proximity) forming the building blocks for short term and working memory: simple chunks represent inferences about what ought to occur together in time or space based on past experience. The predictions implied by these chunks about lower order perceptions (e.g. contours) may be

[26] As in inference process, this can be formally connected to the "free energy" principle outline in the section on embodied representation.

[27] Quotes in parentheses indicate the imagined hypotheses and motor commands of a problem solver.

[28] Ballard et al (1997) have suggested a special role for saccades occurring at time scales of 0.3 seconds, suggesting that they play a special role in embodied computation.

responsible for "filling in" perceived gaps in a given stimulus (for example, the illusory contours seen in the Kanisza triangle (appendix B) might arise from a predictions derived from chunked (prior, learned) information about contrast and the corners of shapes). At the "long" time scale at which most problem representations evolve (one long enough to incorporate multiple problem solving attempts and consequent feedback), the problem representation forms a 'theory' of the problem[29]. Moves are sequential tests of hypotheses (derived from the problem representation) about probable sequences leading to the goal state. A solved problem represents a confirmation of the theory (correct problem representation) while a persistently unsolved[30] problem may necessitate a revision of the theory of the problem (problem restructuring)[31].

To briefly digress (in order to connect the description of problem solving above to the class of relevant "free energy minimizing" biological and computational models

---

[29] Problem representation is composed of perceptual experience and prior knowledge. Perceptual information results from the process of inferring (and testing) underlying causes of sensory stimuli. Prior knowledge can be argued to be the product of past rounds of hypothesis formation and testing in circumstances similar to those of the problem the subject is engaged in solving.

[30] "Persistently unsolved" might suggest that subjects include estimates of problem difficulty and subjective skill level in their problem representation to form an estimate of time-required to solve the problem. This suggests a time scale for persistence on the problem, and/or a time-scale after which restructuring is more likely to occur. It may be that this time scale corresponds to the amount of time necessary to reach impasse.

[31] It should be noted that within the problem solving process there are multiple time scales at which inference processes may unfold: For example, selecting a sub-goal (an inference as to the sub-goal most likely to reach the overall goal) and then picking a move to reach that sub-goal (inferring the move most likely to reach that sub-goal). The relevant time-scales here are the scale of time taken for multiple moves (the expected number of moves needed to reach the sub-goal scale), and time taken for a single move (move scale).

(explored at length in appendix F)[32]), the above description may imply that subjects' inferential processes are hierarchically organized, where information integrated over longer time scales generates expectations about the behavior of phenomena at shorter time-scales: The choice of problem representation, derived from prior knowledge (longest time-scale) conditions expectations about goals. Goals condition expected sub-goals (intermediate time-scale (multiple moves)), and the choice of sub-goals generates expectations about the next best move (short timescale). Perceptual processes, themselves similarly hierarchically organized (Lee & Mumford, 2003; Friston et al. 2012; see also the introduction), constitute the shortest time-scales.

In a related bottom-up process, information from shorter time-scales alters top-down expectations provided by computations at (and about) longer time-scales. For example, failure to reach a particular sub-goal (time-scale: individual moves), repeated multiple times (longer time scale) may lead problem solvers to adopt a new sub-goal (top-down change). Since top-down processes (information about phenomena at long time-scales) attempt to predict lower level (faster-evolving) short time-scale phenomena, feedback provided by lower level processes can be used to update high-level predictions[33]. The mismatch between a prediction derived from a higher model level (inference about the world at a longer time scale) about a lower level's state, and the

---

[32] This refers to models of predictive coding; Predictive coding is an anatomically and physiologically realizable, mathematically tractable and optimal, and (somewhat) empirically verified Bayesian modeling framework for understanding hierarchically organized brain/behavioral processes.

[33] For neuro-biological account of this process see Keibel, Daunizeau, and Friston (2008) or Friston and Keibel (2009).

actual state at that level, called prediction error, is used to update future high level predictions[34]. The Bayesian process by which slowly evolving (long time-scale) system states are used to derive predictions for fast-changing (short time-scale) system states, and which uses the prediction error between the expected states of the system at lower levels (shorter time-scales) derived from expectations at longer timescales, and actual system states at those lower levels, to update higher level (longer-time scale) predictions, is called predictive coding . Predictive coding has been used in numerous treatments of brain and behavior, as a principled Bayesian approach for anatomically realistic computational models of perceptual (and other) learning (Creutzig & Sprekeler, 2008; Friston & Keibel, 2009; Huang & Rao, 2011; Rao & Ballard, 1999; Rao & Ballard, 2004).

Understanding problem solving as a process of predictive coding elaborates and improves upon the definition, offered previously, of problem solving as a hierarchical inference process over multiple temporal and spatial scales. Three additional factors favor a predictive coding account of problem solving: Bayes optimality (Friston & Keibel, 2009) and efficient coding,[35] (Huang & Rao, 2011), converging evidence of

[34] A qualitative example of prediction error in problem solving: Failure to achieve a sub-goal after many attempts might lead problem solvers to believe that the sub-goal is actually an impediment to reaching their goal (accumulation of prediction error, where the initial prediction was the probability of success given the original sub-goal), and cause them to instead infer that a different sub-goal is a more probable candidate for successfully reaching the goal-state).

[35] Reducing the redundancy of a signal by capturing only its most informative elements.

biological implementation[36] of predictive coding across multiple spatial and temporal hierarchies in the brain[37] (Huang & Rao, 2011; Lerner et al 2011), and correspondence and compatibility with older models of conceptual change and problem solving. Of particular interest for the 9-dot problem, gestalt principles can be understood within a predictive coding framework (Van de Cruys & Von Wagemans, 2011; Fang, Kersten & Murray, 2008)[38]. From an embodied perspective, predictive coding models of problem solving have the additional advantage of more easily specifying how, exactly, motor movements are coupled with sensation and perception (movements may occur to adjust

---

[36] Additional evolutionary arguments favor predictive coding. Predictive coding provides adaptive value by predicting salient events, and altering movements in line with those predictions. More-over, since processing delays are inevitable when performing computations, any adaptive system, acted on by selection pressures, transforming inputs into a set of useful outputs, must necessarily be predictive in order to compensate for processing delays. Consider the visual system; at its most basic it must at least overcome processing delays inherent in coordinating its inputs, therefore to be of use, any computation making use of light that has hit the retina, should not try and deduce what "was" the case when the computation started (when light hit the retina), but what "will be" the case when the computation is finished. See Changizi and associates (2008a; 2008b) for more on "Predicting the Present". Predictive coding may be the mechanism by which "predicting the present", takes place.

[37] Levels of the predictive coding hierarchy are said to correspond to structural and functional segregation in the brain.

[38] Gestalt filling-in and other phenomena arise from top down predictions of what "ought to be" in an image given prior knowledge and current stimuli. When stimuli imply a particular representation strongly enough (even though the implication may be incomplete) prior knowledge (in the form of top down predictions) fills in the gaps left by the 'incomplete' stimuli. Further, the hierarchical nature of the visual system (both functionally and anatomically (see Murray et al (2002) for a simple example) is conducive to predictive coding explanations; it may be that various gestalt rules (e.g. figure-ground relations, proximity, co-linearity, etc.) map onto functional segregations in the visual system, and thus map onto separate levels of a predictive coding hierarchy. Though there is some room for competing interpretations of evidence for this, see De-Wit, Kubilius, Wagemans, & de Beeck, (2012).

environmental stimuli to conform with internally generated expectations about the environment) (Friston et al. 2012).

Predictive coding can be interpreted with the vocabulary of restructuring. Each level of a predictive coding hierarchy can correspond to set of 'like' elements either tightly or loosely chunked (these elements might range from early visual primitives (e.g. contours), up to networked semantic relationships). Each chunk[39] (level) is both a superordinate chunk, a particular macro-organization of lower level representational chunks[40], and a subordinate chunk, a possible building block for still larger chunks. Levels of the hierarchy interact in such a way that ascending interactions (from lower to higher levels) influence the tightness of superordinate chunks, while descending interactions (from higher level chunks to lower level chunks) act to constrain the possible

[39] The identification of separate neural correlates for chunks of various levels (perceptual vs. semantic) in insight, as well as the suppression of neural activity at one or the other of these levels during chunk decomposition (Luo, Niki & Knoblich, 2006; Wu, Knoblich, Wei & Luo 2009; Wu, Knoblich, and Luo 2012; Wu, Wu, & Luo 2010), suggests that the anatomical machinery needed to distinguish between hierarchical chunk levels is in place. If anatomical location also implies a separation of time-scales, as suggested by Keibel and Friston (2008), the evidence cited above would imply that the process of chunk decomposition occurs in a hierarchical manner which might be conducive to (or even indicative of) predictive coding.

[40] Or, perhaps, a probability distribution over all macro-states contained in that chunk.

arrangements of subordinate chunks[41] [42]**.** Loosening a higher order chunk relaxes constraints on lower order chunks, while tightening a higher order chunk may impose constraints on the chunks below[43]. This is similar to the function of prediction in higher levels of predictive coding: high level predictions, by acting as Bayesian priors for lower order predictions, can constrain the space of possible low level predictions. Chunk

---

[41] In more Bayesian terms: The likelihood of any two mid-level representational chunks to be associated with one another (bound more tightly together) might be specified by shared connections from lower level chunks (e.g. the number of lower level chunks that two mid-level chunks have in common). Whereas, the probability of any two mid level chunks being part of a higher level chunk (composed of more than two elements_ is related to the probability of that chunk having occurred in the past.

[42] A useful example: In the famous image/perceptual-illusion "my wife or my mother in law" (Appendix C), the viewer recognizes either a picture of a young woman, or an old lady, but never both (or some hybrid) at the same time. In terms of the explanation given here the two top-down representations constrain the possible interpretations of the arrangements of lines and edges in the picture. The priors (based on past experience) for the hypothesis "this image is a young woman", and this "this image is an old woman", are stronger than those for "this is an image of an old and young woman at the same time" or "this picture is neither an old woman or young woman, but something else entirely". Bottom up prediction error, accumulating over time, forces the oscillation between the two dominant percepts (young v. old woman). (See Howy et. al (2008) for further explanation).
In the chunk interpretation, prior experiences constrains the possible macro-states (arrangements) of the many possible underlying subordinate chunks (contours) into just two probable superordinate "meaningful" chunks (young vx. old woman). A top-down interpretation, once selected further meaningfully sub-chunks each element of the picture (e.g. the ear of the young woman is the eye of the old woman). From a bottom up perspective, changing certain features, peripheral to one interpretation (e.g. young woman), but more central to the other (e.g. old woman), such as deleting the young-woman's necklace (old woman's mouth) and covering her ear (old woman's eye), drastically increases the likelihood of seeing the young woman (and decreases, to almost zero, the likelihood of seeing the old woman) in a manner disproportionate to the mere number of dots removed.

[43] This is similar to Knoblich et al's (1999) discussion of tight vs. loose chunks and hierarchies of constraints, however in the proposed predictive coding interpretation, chunks themselves imply or relax constraints.

decomposition, driven by the dissociation of chunk-elements based on experience may correspond to the altering of high-level predictions by bottom up prediction error in predictive coding. Biologically, these interactions may play out along the lines of Barsalou's (1999, 2009) (predictive) perceptual symbol systems, perhaps implemented by attractor network dynamics in the brain (for example: Akrami, Russo, & Treves, 2012; Lerner, Bentin, & Shriki, 2012A & B; Russo & Treves, 2013; Treves, 2005).

More broadly, predictive coding captures the insight of Knoblich et. al (1999) that "knowledge structures tend to undergo local, peripheral, or superficial changes before they undergo global, central, or fundamental changes" (Knoblich further cites Chi (1992), Vosnaidou, (1994) and Rokeach (1970) [44]). Knoblich's local, superficial changes (roughly) correspond to the fast-changing, short-time-scale phenomena at the bottom of the predictive coding hierarchy, while Knoblich's central and fundamental features correspond to slow changing (long-time scale) phenomena nearer the top of the predictive coding hierarchy.

If the hierarchical inference process characterizing embodied problem solving is implemented using predictive coding, how is the nine-dot problem to be understood with predictive coding? Or rather, what does the embodied perception approach, viewed through the lens predictive coding, imply that other theories do not? By connecting action, perception, and understanding across multiple cognitive scales, the predictive coding explanation unifies the nine-dot's multiple sources of difficulty in one explanatory

---

[44] Imre Lakatos' conception of scientific research programmes (Lakatos, 1975), with core and auxillary hypothesis may also broadly fit this description.

framework. Moreover, the predictive coding account offers an explanation of the interaction between hints and impasse, and generally suggests useful connections between timescales, embodied processes, and insight (of which the relationship between hints and impasse is an example). Finally, the predictive coding account suggests a general explanation for the qualitative differences between insight problems and non-insight problems, in line with Stephen and Dixon's (2009) constraint breaking/self-organization explanation for insight: the accumulation of prediction error at various levels of the perceptual/cognitive hierarchy eventually necessitates higher level "explanations", which (may) correspond to constraint-breaking and reforming in Stephen-and Dixon's model (Friston, Breakspear & Deco, 2012; Stephen & Dixon 2009;). (It might be hypothesized that non-insight problems only require solvers to adjust prediction error locally (relative to a given model level), without appeal to further hierarchical levels.)

As mentioned previously, difficulty in the nine-dot problem may arise from multiple sources. Kershaw (2002, 2004) characterized these as perceptual sources, prior knowledge sources, and process sources. In the predictive coding model, these sources map onto discrete levels of the coding hierarchy. The lowest level of the hierarchy consists of perceptual factors. The processing time-scale of this level is roughly equivalent to the time-scale of meaningful saccades, described by Ballard et al (1997) as either the time-scale of attention (50msec) or the time-scale of embodiment (0.3 sec)[45].

---

[45] The time scale at which these various processes play out is important for experimental purposes, as it suggests the necessary "grain" of analysis needed to capture events at this level of the cognitive hierarchy. Theoretically, these timescales are also of interest, as they further imply the hierarchical organization of cognition. (See Ballard & Rao, (1997) for more).

The second level of the hierarchy is composed of process factors, such as those identified by Macgregor et al (2001). Since Macgregor identifies look-ahead and working memory as key process elements, the time-scale at the process factor stage is (hypothesized to be) equivalent to Ballard's cognitive time-scale of 2-3 seconds. A third and fourth process levels may also be at work, though at present these scales are harder to define. The time-scale at the third level is the time-scale necessary to make at least four moves (definitively ending in either success or failure, providing the subject feedback on the effectiveness of a strategy). The fourth possible time-scale (varying by individual), is the time-scale at which subjects stop actively engaging with the problem after a series of failures. This may correspond to the period at which impasse begins, and at which hints are most helpful (Moss, Kotovsky, & Cagan, 2011)). The final source of difficulty, prior knowledge, constitutes a fifth level of subjects' hierarchical model. It contains both tacit (e.g. past experience with "connect the dots") and explicit knowledge (e.g. semantic relationships for interpreting rules) stored in long-term memory. Its time-scale is the indefinite past.

The ways in which sources of problem difficulty interact are many and varied. For the sake of brevity, this paragraph and the next will analyze only some of the possible top down and bottom-up contributions to Macgregor et al's information processing model of the nine-dot problem. Macgregor's model is chosen because in the hierarchical framework suggested here, Macgregor's information processing level constitutes an intermediate level in a hierarchy, midway between perceptual factors and long-term knowledge factors, and is therefore a useful point from which to analyze the interaction of top-down and bottom connections. (Further, Macgregor's model is predictively valid).

Macgregor's model takes various perceptual elements as primitives. However, these perceptual elements are not static or fixed, and independently contribute to problem difficulty (Kershaw, 2000; and Kershaw & Ohlsson, 2004). A more complete model ought to account for changes in perceptual elements. For example, Macgregor's model defines both "adjacency" and "line" a priori, noting that the model's definition of "line" is consistent with Gestalt principles. However, precisely how adjacency is used to group elements (to form a good gestalt) is defined by the principles of the visual system. In fact, adjacency along with proximity, as gestalt factors, may be considered lower-level factors contributing to a higher-level definition of "line" (or contour) as demonstrated by Claessens and Wagemans (2008) (see also Elder and Goldberg (2002)). Altering the perception of lines that imply problem boundaries from the bottom up, by the placement of a additional dots (thereby fostering different Gestalts), including perceptual (and motor) practice with figures in the shape of the intended solution, or providing mixed training between figure-ground grids and regular instances of dot problems (Kershaw 2002)) can effect problem representation on the perceptual level level, by, for example, increasing the number of allowable lines in problem solvers' repertoire, or relaxing constraints related to perceived problem boundaries (the implied edges of the box)[46].

From a top down perspective, Macgregor's "hill climbing", or difference reduction strategy, if extended to cover multiple trials and including a weighting factor (used to discount lines involved in failed solution attempts) can be interpreted in terms of prediction error reduction: The (high level) goal state (four lines, covering all dots), prior trials in the problem, and Macgregor's value function (2001) are used to 'predict' the most optimal course of moves. Past experience might be used to weight moves in terms of their probability of leading to a successful solution. (In terms of predictive coding, the relevant time scale for this level is the length of time taken for prior trials). Moves are executed, and feedback ("was the goal reached or not?") may be used to calculate a prediction error function, which is used to re-weight the moves used in the solution attempt. Since moves are probabilistically selected based on weight, re-weighting serves to update predictions about effective move sequences. In this case, the top-down features (goal state, prior experience as reflected in the current weights of moves) at the time scale of multiple trials, are used to predict (select) a (probably) successful move sequence, while bottom up prediction error (at the time-scale of a single trial) is used to adjust the higher-level predictions (move weights).

---

[46] Even Maier and Casselman's finding that drawing problem solvers' attention to the implied box, (by drawing a line around the 9 dots and telling solvers to work outside it), can be considered a (failed) perceptual manipulation albeit one that, perhaps, increases difficulty. By drawing attention to the implied box, they may have reinforced the bottom up perception of the dots as forming a box, in a manner that was difficult for higher level cognition to over-ride.

The examples above are but a small portion of the many interactions between levels in a predictive coding interpretation of the nine-dot problem, for an example of (hypothesized) hierarchical interactions in the nine-dot problem in the empirical literature, the reader is directed to the review of Ollinger et al.'s (2013) work in the literature review (p.24). For the purposes of this paper, the primary interactions of interest are those reflecting embodied processing and coupled (boundary cross) hierarchical processing. Since time-scale seems to be an important factor in understanding embodied cognition (Ballard et al, 1997), and time-scales are also crucial for understanding the dynamics of predictive coding, it is helpful to specify exactly "what" occurs "when" in the hypothesized predictive hierarchy of the nine-dot problem.

The time-scales suggested by Ballard seem to suggest the "grain" of at least three levels of a predictive coding model: attention (50msec), embodiment (.3 sec), and cognition (2-3 sec). The embodied time scale is defined by the time it takes to compute/ process a sensory input and then act on the results of the computation. For this paper, the content at these differing levels is also hypothesized to be hierarchically organized. The shortest time-scales use the most basic environmental stimuli to process basic perceptual factors (e.g. contrast, distance, and adjacency), while slightly higher levels use the inputs from the basic perceptual factors to form predictions about contours, and still higher levels produce a fine-grained analysis of shape. At the cognitive time-scale (2-3 seconds) lower level perceptual factors combine with working memory to influence the selection of possible moves. Further, a long-time scale, relating to multiple trials (and failures) in nine-dot solving, is hypothesized to be important, connecting the prediction-error over-multiple-trials account from the previous paragraph to Moss et al's (2011) explanation for the effects of the timing of hints[47] and the onset of insight.

Broadly, the goal of sketching a theory of embodied problem solving that connects embodied behaviors to high-level cognitive processes, is, for the purposes of this paper, complete. A general predictive coding account of problem solving has been combined with evidence for the time-scales of embodied cognition to provide a principled, biologically and cognitively plausible explanation for embodied problem solving and problem representation. This theory, and the nine-dot specific elements described above, are used below to develop an experiment examining the role of embodied factors in the nine dot problem.

The suggested experiment uses directed eye movements to manipulate representational change in the nine-dot problem. It combines hints at multiple levels of the hypothesized cognitive and perceptual hierarchy with eye movements and fixations at time scales that reflect Ballard's division of time scales into attention, embodiment, and cognition. It further tests predictions about long time scales by varying the timing of the hints.

More specifically, embodied perceptual hints, in the form of a directed eye movement task, are given after participants reach impasse. The directed eye movement task requires participants to attend to single stimuli appearing sequentially on an otherwise blank screen. The patterns implied by the order of the stimuli's appearance are

[47] The hypothesis here is that the (precision weighted) prediction error is considered a driver of attention (Feldman and Friston, 2010), which might facilitate the "noticing" phenomena necessary to find a hint useful.

designed to either facilitate or hinder performance on the nine-dot problem at different perceptual levels: A random sequence of dots is expected to have no effect on performance. A sequence of dots whose spacing is governed by a power law, similar to that used for combining elements into contours (Claessans & Wagemans, 2008; Elder & Goldberg, 2002), might either reinforce existing perceptual biases, or help decompose the (very low level) perceptual chunk of implied contours, depending on the power law's exponent (relatively large (negative) exponents should help break perceptual chunks, while small (negative) exponents might reinforce them). Dots sequentially appearing to form the perimeter of a square (the 'shape' level of the perceptual hierarchy) will likely inhibit performance.

Dots appearing as the corners of, and in the order of moves made in, the correct solution sequence should facilitate performance. Dots whose sequential appearance does not trace the correct solution sequence but does form the overall solution shape (e.g. starting in the wrong location but then tracing the correct shape), are used to test the interaction between perceptual factors and working memory/process factors. Chunking the dot sequences into the solution shape should facilitate performance (indicating the strength of perceptual factors over process factors at this level), whereas viewing the dot sequence as the suggested order of moves (dots as independent elements, rather than part of a whole), a process factor, will inhibit performance. At this level, there should also be an interaction between the duration a dot is viewed and the type of processing the viewer engages in. Viewing each dot for 0.3 seconds (the embodied level) leads to a total viewing time for 4 dots of 1.3 seconds, closer to the cognitive level, suggesting that processing at the embodied level might lead to the formation of a useful perceptual chunk. Viewing each dot for 2-3 seconds each might lead to fixation on individual dots or positions, leading to dots to be seen only as parts of moves, inhibiting overall performance on correct-shape, wrong-order hint variant.

Dots whose sequential appearance is the solution shape, but further extend the implied fourth line by adding an additional dot to the sequence, so that the implied fourth line extends beyond the implied boundaries of the square formed by the hint's three initial dots (in a manner similar to farthest dot in Macgregor's thirteen dot variant) test the interaction between perception and output. Subjects receiving the "extended" dot sequence may perform better than the non-extended hint (the implied line clearly breaks the perceptual gestalt of any implied square, in either by problem stimuli or the order of hint-dots). Even if performance is the same, comparing the length of the fourth line drawn by extended vs. non-extended hint conditions would suggest interactions between perceptual factors and outputs. Similar performance, but different final, participant-drawn, line lengths would suggest that some perceptual factors influence solution outcomes at a low level (motor skills), while not impacting process level considerations.

A third set of independent variables, concerning time-scales, applies to the length of time a dot appears on the screen, and thus, the length of time a subject fixates on that dot. Time-scales are divided into short, medium, and long time-scales according to Ballard's hierarchy (short, attention (50(msc), medium, embodied (0.3 sec), and long , cognition (2-3 sec)). Each perceptual manipulation then, has three variants (short, medium, and long). It is expected that the short variants will have less effect (either facilitating or hindering) than medium (embodied) variants, whereas the difference between medium (embodied) and long (cognitive) variants is more situation dependent. If long variants are congruent with a successful high-level strategy implied by the sequence of dots (e.g. sequence forms solution shape, and starts in the right place), they should improve performance over the embodied level. If long variants are lead to interpretations of dot sequences that hinder solutions (e.g. sequence forms correct solution shape, but has the wrong order), but medium-lenght (embodied) variants suggest an interpretation that is more in line with the solution sequence, then medium variants will lead to better performance.

These predictions reflect a general hypothesis that task performance is facilitated by perceptual and embodied hints that more easily map onto the problem's eventual solution, and that the effectiveness of hints depends in part on matching the perceptual or cognitive level of the hint to the time-scale appropriate for that level, where embodied processes serve to bridge low-level (perceptual) and high-level (cognitive) processes through action (in this case, eye movement).

Appendix H: Example of nonlinear top-down and bottom up interactions in the nine-dot problem, using Ollinger et al.'s (2013) and Chronicle's (2001) results.

Along with Macgregor (2001) and Kershaw and Ohlsson (2004), Ollinger and colleagues (2013) have gone the farthest in providing a thorough representational change account for the nine-dot problem. Ollinger et al's results are particularly informative as to the interaction of problem elements in problem representation. By placing arrows on the dots implicitly indicating solution paths, Ollinger and colleagues caused participants to widen their initial (wrongly constrained) search space. Adding information to indicate line intersection and blank spaces to indicate non-dot turns drastically improved performance (from 11.76%, 26.47%, 44.12%, 81.82%, respectively. 11.6% is the baseline solution rate). Ollinger et. al's framework supports an information processing understanding of the problem (search space first broadened, then constrained by higher level semantic factors (arrows)), but adds a perceptual element (the blank space for non-dot turns), which appears to trigger the release of yet another constraint, the non-dot turn constraint.

By adding blank dots on the edges of the box, as in Chronicle (2001), Ollinger et al. changed the perceptual representation of the figure. Unlike Chronicle, however, this representation did not seem to imply a figure-ground relationship between the additional dots and the box, but rather served to highlight a possible location for the continuation of a (solution) line. In doing so Ollinger's elaborated "box and extended corners" object allowed participants to relax the non-dot constraint.

Ollinger et al.'s account is an excellent demonstration of the interaction of hierarchical factors in problem restructuring: Semantic content (in the form of arrows) helps to relax lower order constraints (where and how to search the box), which were initially implied (in part) by perceptual factors. Highlighting those perceptual factors *after* the top down constraints were relaxed allowed participants to successfully re-interpret the bottom up clue provided by the blank space. Without the higher-level constraint relaxation, participants could not (as demonstrated by Chronicle et. al (2001)) effectively elaborate what they perceived to be a contrasting-edge into an indicator of a possible line extension/location for a (non-dot) turn. The semantic hint created the possibility to relax perceptual constraints, making optimal information processing (search) possible. With the semantic (arrows) hint alone though, solution was less likely, because the implied constraints from past experience and rule interpretation, "no non-dot turns", had not been relaxed. Only by perceptually elaborating the structure by adding additional dots (a bottom-up perceptual manipulation) was the last non-dot turn constraint (somewhat) circumvented (whether or not it is fully relaxed is unclear, perhaps there was some further auxiliary, re-interpretation process as well).

In Ollinger's work, the bottom up perceptual trigger (the additional dot) is quite helpful when paired with additional hints, while in Chronicle's (2001)[48] the same dots seemed useless or worse. In this case, problem representation, understood to include the contextual whole formed by the problem elements, seemed to independently contribute to

---

[48] This sort of non-linear hierarchical interaction is precisely what the model contained in the last section of the paper aims to capture.

problem difficulty, over and above the (possibly expected) additive contribution of the problem's individual elements. That is, not only are the individual elements important, it's how they hang together: a thought that would sound none-to-foreign to the original Gestalt researchers of the nine-dot problem.

Either the last point is indicates deeper structures at work in problem solving, or a spurious coincidence weakly justifying circular and tiresome reasoning (To a hammer all problems look like nails, to a gestaltist all nails look like nine dots and matchsticks). This paper takes the former view.