

Chapter 9

Conclusions and Future Work

I think that the single most important contribution of this research project is the way it sets out the distinction between two contrasting perspectives on concepts and two, corresponding, approaches to understanding them; then demonstrates the way that most contemporary theories of concepts line up on one side of the divide or the other. Are concepts (“mental”) representations, or are they (non-representational) abilities? They must be both. The seeming inconsistency between the two perspectives is nothing more than a reflection of the unavoidable limitations of our conceptual horizons. A version of conceptual spaces theory is, I believe, best placed to bridge the divide and bring the two sides as close together as is practical (and, indeed, either possible or desirable) to do. The principal contributions of each chapter are:

Chapter One: A unique treatment of literal versus metaphorical boundaries as they apply to cognition in general and concepts specifically; a novel argument for the importance of metaphysical biases to the discussions in the remainder of the thesis.

Chapter Two: The distinction between concepts as we reflect upon them and concepts as we possess and employ them non-reflectively – part of what I later (Chapter Five) call the *toggling effect*; a novel treatment of representations, and the relationship between concepts and representations, tied to where one locates the observer.

Chapter Three: The distinction between the private and public aspects of concepts, and an argument why that should be understood in terms of contrasting aspects as opposed to different kinds of concepts; a novel argument for including *evolvability* among the core properties of concepts, and against including *introspectibility* and *articulability*, as commonly understood.

Chapter Four: A novel classification of concepts by types of conceptual agents, types of referents, and manner of use; a novel application of Donald’s four stages of cognitive-cultural evolution to the evolution of concepts, addressing an oft-neglected issue in theories of concepts.

Chapter Five: A novel elucidation of the relevance of self-reference to theories of concepts; the notion of concepts as *necessary fictions*; a novel argument in favour of the extended mind hypothesis that avoids the risk of cognitive bloat; a novel response to Penrose’s argument that human understanding is not bound by Gödel’s Incompleteness Theorem.

Chapter Six: The unified conceptual space theory as an extension of conceptual spaces theory, showing how all of an agent's many conceptual spaces come together into a single space of spaces; a clarification of the relationship between conceptual spaces theory and enactive philosophy.

Chapter Seven: A detailed theoretical account of the co-emergence of concepts and experience, as an application of conceptual spaces theory and the unified conceptual space theory; the notion of *concepts as expectations*, increasingly structuring the experience that structures them.

Chapter Eight: A software implementation of many aspects of the unified conceptual space theory and illustration of the benefits of such an application to the further iterative development of the theory; at the same time, a visually and theoretically distinctive approach to mind mapping and non-linear structuring of ideas.

9.1 Looking Back

It might be useful at this point to consider the research proposal I wrote five years ago, as I was embarking on this project¹.

I see my thesis being a specification of a KR formalism and the creation of a toy environment as a springboard for philosophical discussion. The purpose of such a formalism can be left intentionally ambiguous, at least initially: are we merely representing the conceptual knowledge of agents or are we representing the way that knowledge is represented by the agents themselves? We can simply say: this is a way of representing conceptual knowledge (or some appropriate sub-domain?) with a uniform representation; let's see what the consequences are. Areas for philosophical exploration:

- What is the essential nature of "concept"? Though representation seems inevitable, the view of conceptual knowledge as a set of representations is not. What is the value for suggesting a uniform representation for concepts? Is a truly uniform representation (a) formally describable and (b) even possible?
- We have our notions of "real world" (or "external world") and "mental world", and in popular discourse we generally treat them as being quite separate and distinct. But medical and psychological research would seem to indicate that there is never any uninterpreted ("unmediated") translation of information from the external world to the mental one. So: what is the nature of the relationship between external world "object" and mental world "model"? Are they necessarily (a) separate and distinct; (b) related, possibly even isomorphic; or (c) simply different levels of the same thing (e.g., "model", "model of the model", "model of the model of the model", and so on).
- Human language is, of course, another form of representation. We can say that we use it to mediate between our own mental worlds and our own models of the external world on the one hand, and other people's mental worlds and their models of the external world on the other. Linguists at least since Chomsky have postulated a common structure to all human language; would this structure be related in any way to how the mind itself structures knowledge, chosen specifically because of that similarity?

¹This version has been shortened from the original.

- If we take as starting point (for the sake of argument) that "mind" consists of some set of representations, what (if any) is the usefulness of viewing "consciousness" as just another representation, but one that is unique insofar as being a representation of the system itself (thereby creating what Hofstadter has been fond of calling "tangled loops")?

Looking more closely at the nature of "concept", what understanding am I starting from? Premises:

- Considering concepts as mental representations: "... Every aspect of thinking can be viewed as a high-level description of a system which, on a low level, is governed by simple, even formal, rules" (Hofstadter, 2000, p. 559).
- As a general rule, concepts can be viewed as an "atom" or broken down into smaller concepts. That is to say, any concept has certain requisite components. (My concept of "man" requires a "head" and a "torso"; otherwise, it's not a "man". Other components, such as two arms and two legs, are attached to my concept of "man" but not requisite.)
- At some point, this decomposition has to stop. One can postulate an atomic "concept" from which one will derive all other concepts. This is the "uniform representation" suggested above: the underlying "building block". (Philosophers probably would not want to call such a low-level construct a "concept" at all.)
- Any concept has certain requisite descriptive properties or parameters. (My concept of "man" requires "man" to be "alive"; otherwise the appropriate concept is something different, like "corpse".)
- Any concept exists within a certain context. Any concept that has no context has no meaning.

Most of the themes of the present work are already reflected here: the relationship between concepts and representations, concepts and language, concepts and proto-concepts, concepts and referents, concepts and context; the search for a uniform representation formalism; the questions of conceptual ontologies, conceptual properties, and metaphysical perspectives; the plans for a software-based test application.

If I were to criticize myself, it would be for attempting to do what I claimed in Chapter 5 I could not: provide a too-complete theory that tries too hard to be universal. More practically, I would criticize myself for allowing the philosophy of mind to play too dominant a role in the final product and failing to maintain a balance between largely armchair-based theorizing and hands-on application: one of the very tensions I am claiming to be most important to hold onto (Section 6.1.3). AI researchers will not see the immediate relevance to how they do "knowledge representation" – the thesis offers no neat module that can simply be slotted into their work; cognitive scientists, unless they are already in the grip of Fodor's angst (Section 1.3), probably not see the motivation for making implicit assumptions about concepts explicit. Meanwhile, the debate between concepts-as-(mental)-representations and concepts-as-abilities will go on.

As noted at several points, the present account makes no real effort to define salience. Neither does it address conceptual dynamics in any satisfactory way: as noted in Section 6.2.3, a mature unified conceptual space should be able to capture a rich notion of dynamics internally: i.e., it should be able to capture something about the way it and its world model change.

The argument on metaphysics and boundaries was recently accepted for publication (Parthemore, 2011), much to my satisfaction. Chapters Two through Four present, perhaps, the most polished material, having been re-written any number of times since the first draft of what was to become all three chapters was written almost exactly five years ago; but the argument in Chapter Five, for all that it has gone through several versions itself, still is not so clear as it needs to be. Chapter Six, though it represents published material (Parthemore and Morse, 2010), sets out the framework for the unified conceptual space theory neither to my own nor Peter Gärdenfors' satisfaction; instead, he says "it's getting there". Chapter Eight is much briefer than I would have cared for, and the software program itself is, though visually quite suggestive, still some distance from being put to practical use. That said, it has usefully revealed a number of areas in which the unified conceptual space theory needs to be pressed forward if a truly unified space of all the conceptual spaces is to be achieved.

9.2 Looking Ahead

People are willing to rank simple objects of different shape and colour on the basis of "similarity". If machines are to reason about structure, this comparison process must be formalized (Aisbett and Gibbon, 1994, p. 143).

Like all theses, this thesis is a work in progress. The notions of *concepts as necessary fictions* (Section 5.2.3), *the toggling effect* (Section 5.4), and *concepts as expectations* (Section 7.3.1) in particular could all use further elucidation, as could the public/private distinction (Section 3.3.3). The "response to Penrose" (Appendix A) could use further polishing and, if it continues to stand the test of argument (and is sufficiently distinct from McCullough's own argument (McCullough, 1995)), submission for publication.

To my mind, the most exciting avenue for further research lies with development of the test application presented in Chapter Eight. Just because theories of concepts are not amenable to direct empirical testing does not mean that they cannot be tested in a variety of extremely productive *indirect* ways. Empirical testing based on "Charley" could include:

- Exploring how natural (or unnatural) naive users find the application, as (indirect) evidence for some isomorphic relation between how it structures information and how their underlying conceptual thought processes do so.
- Exploring how naive users make sense (or fail to make sense) of a map created by someone else for a particular domain: how they "correct" it, etc.
- Exploring how experts judge maps created by novices.
- Comparing how different users (experts and novices) carve up a particular domain.
- More broadly: discovering shortcomings in either conceptual spaces theory or the unified conceptual space theory.
- More ambitiously: automatically extracting conceptual dimensions from a suitable corpus using some version of *latent semantic analysis* (LSA)².

²For a good introduction to LSA, see (Landauer et al., 1998).

- Most ambitiously: exploring how, *per* the brief remarks at the end of Chapter Eight, the application could be given some degree of autonomy, e.g. by using a robotic platform, and so close the circular causal loop presented in Chapter Seven. Such an autonomous system would need some way of automatically extracting conceptual dimensions according to some measure of salience and some way of automatically generating the metric for the resulting spaces in a way such as Aisbett and Gibbon have suggested. (See the discussion of metrics in the introduction to Chapter Six.)

That said, before such an autonomous platform can even begin to be explored, the theory must be more developed and stabilized. As in so many other application areas, the top-down-driven approach provided by the mind-mapping application must come first, in order to constrain meaningfully the subsequent search space.

As I set out in the introductory chapter, a primary motivation for this thesis was the conviction that theories of concepts are both central to cognitive science research and much if not most of the time left implicit. I would hope my thesis can provide some motivation in cognitive science not so much toward adopting *my* theory of concepts as explicitly acknowledging *some* theory of concepts. Concepts (as the structuring units of [essentially] all of our structured thoughts) are foundational, and foundations matter!