What is the Role of Conceptual Analysis in Cognitive Science?

Liam C. Kavanagh* (lkavanag@ucsd.edu)
Department of Psychology, UC San Diego
9500 Gilman Drive La Jolla, CA

Christopher L. Suhler* (csuhler@ucsd.edu)
Department of Philosophy, UC San Diego
9500 Gilman Drive La Jolla, CA

* Both authors contributed equally to this paper.

Abstract

Cognitive scientists sometimes find themselves embroiled in debates over the precise definitions of high-level concepts in their fields – COGNITION, EMOTION, SENSE, and so on. The idea behind these debates seems to be that achieving a precise definition of these concepts will be a boon to scientific inquiry. We argue that these efforts of conceptual analysis would benefit from greater appreciation of the importance of such high-level concepts in supporting association or semantic priming, as opposed to deduction. In this associative role, they provide the basis for making connections between related concepts, connections that can then be explored by empirical methods, which in turn yield more precise, but often quite novel, concepts. In combination with well-established work in cognitive psychology on the non-classical structure of natural concepts, this perspective suggests that researchers should be cautious about investing substantial time and energy in attempts to precisely define concepts like COGNITION.

Keywords: concepts & categories; philosophical issues; philosophy of science

Introduction and Background

The investigation of concepts has been a central part of intellectual inquiry since at least the time of Socrates. Today, conceptual analysis remains a cornerstone of academic philosophy. Cognitive psychology, too, has taken a keen interest in the meaning of concepts; inquiries here have often taken the approach of first investigating "simpler" and more tangible object concepts, in the hopes of working their way up to a grasp of more abstract concepts. Despite the substantial effort invested in precise accounts of concepts' meanings, the results have been uneven and at times frustrating. In particular, practitioners of conceptual analysis often doggedly pursue efforts to arrive at precise definitions of high-level concepts, such as COGNITION, EMOTION, SENSE, and so on with less stellar results than are arrived by investigating lower-level concepts such as SMILE, SYSTEM, or CIRCUIT.

We will make frequent use of COGNITION as an example of a high-level concept, but the same points apply, *mutatis mutandis*, to other high-level concepts. We also readily admit that the distinction between "high-level" and "low-level" concepts is, itself, not perfectly precise, but do propose the following as a general marker of the distinction: low-level concepts are those which tend to play greater roles in mechanistic accounts of empirical results, while high-level concepts are those that do not (for example, we rarely, if ever, would explain some empirical result simply by saying that it is "cognitive").

We argue that these efforts of conceptual analysis would benefit from greater appreciation of the importance of such high-level concepts in supporting association or semantic priming, as opposed to deduction. In this associative role, they provide the basis for making connections between related concepts, connections that can then be explored by empirical methods, which in turn yield more precise, but often quite novel, concepts In combination with well-established work in cognitive psychology on the structure of natural concepts, this perspective suggests that researchers should be cautious about investing substantial time and energy in attempts to define concepts like COGNITION.

The Allure of Precise Definitions

There is something about precise conceptual definitions – definitions which lay out the characteristics that all members of a category *must have* – that continues to attract our curiosity. This is despite the fact that introductory cognitive psychology texts have for several decades stated plainly that natural concepts almost never take such a form. Persistent targets of definition-seeking include concepts such as COGNITION (e.g. Adams & Garrison, 2013), EMOTION (e.g. Mulligan & Scherer, 2012), and SENSE (Keeley, 2002).

We believe that there are two major assumptions, often tacit, which together explain the continuing pull of precise conceptual definitions. The first assumption is that precise definitions are out there waiting to be found, if only we look (and think and argue) hard enough – or at least that there's a good enough chance that such definitions are out there to make the pursuit worthwhile. The second assumption is that obtaining such precise definitions will be a boon to scientific inquiry and understanding . Both of these assumptions, we will argue, turn out to be dubious. This, in turn, casts doubt on whether rigorously delineating the set of referents of terms such as COGNITION is a good use of researchers' time.

Part of the appeal of precise definition is likely attributable to vestigial influence of the *classical view* of reasoning and concepts (Smith & Medin, 1981), according to which concepts (and their associated categories) are defined in terms of necessary and sufficient conditions. This was the dominant view of concepts for almost the entirety of western thought, with its origins often traced back to Aristotle (Smith, 1997). Concepts of this kind are the epitome of precision; a classical definition of a concept promises precise demarcation of the boundaries of that concept, and allows precise, deductive inferences.

The Classical View of Concepts

The familiar tale of the failure of the classical view of concepts hardly needs detailed recounting, and as such we will be exceedingly brief here. The key point is simply that empirical research has established that concepts usually do not have a unified definition (i.e., a set of features common to all and only members of the concept) or sharply demarcated boundaries – Mervis & Rosch, 1981). This historical record provides strong reason to think that any proposal of classical criteria for a given concept will fail to capture all of the phenomena that intuition demands. Thus, in the case at hand, it is likely that a definition of COGNITION will fail to capture all (and only) those phenomena that we regard as cognitive. Furthermore, even if it gets most of those right, it will fail to recognize degree, since on a classical view of concepts something is either part of the concept (i.e., possesses the necessary and sufficient features) or not (i.e., does not possess the necessary and sufficient features), with no gradations or gray areas.

Debates over the precise criteria for COGNITION are therefore likely to lead to a great deal of ink being needlessly spilled trying to impose a structure that the actual mental concept likely doesn't have. A given definition will include some desirable things, and even where it gets the basic verdict right will fail to recognize gradations of typicality/atypicality. For instance, proposals which assign cognitive status to cellular information processing will be intuitively inadequate if they fail to recognize that these processes, despite possessing features associated with cognition, are not paradigmatically cognitive.[1]

The aforementioned considerations are of course not definitive proof that definitions of cognition are undesirable or impossible. It is not clear that there is any way to prove that definition-seeking for any specific natural concept will, as a matter of necessity, fail, despite the rejection of the classical view by cognitive psychologists and the dubious record of attempts at classical conceptual analysis. We expect that some will continue to hold the intuition that a definition of COGNITION is possible, just as many philosophers hold the very strong intuition that classical analyses of their most cherished concepts are possible despite decades or even centuries of failure (Suhler, in preparation).

Further, even if there is a (satisfactory) precise definition of cognition to be found, history strongly suggests it will be achieved because theoretical *and empirical* inquiry reveals a similar underlying nature of things we take to be cognitive; it will not, in other words, come about through standard methods of *a priori* conceptual analysis (reflection on our folk concepts, thought experiments, etc.). We will develop this point later in the paper by examining the history of conceptual evolution in physics. Before that, however, we want to look more specifically at why, despite the vanishing rarity of classically defined concepts, there exist strong intuitions that concepts like COGNITION admit of classical definition.

Though intuitions are indeed often valuable things, they are not beyond criticism. The most effective response to the belief that we can precisely define concepts like COGNITION, we feel, is to explain how these intuitions in favor of pursuing precise definitions are driven by factors that do not track the actual probability that such definitions are attainable. There are two such factors we wish to highlight here. The first is saliency. The scientific concepts that are likely to be most salient *qua* scientific concepts tend to be those drawn from mature sciences dealing with relatively simple phenomena – viz., physics and physical chemistry (see below). The second is confirmation bias – the tendency to notice the examples that fit with one's preexisting commitments rather than those that do not. In suggesting a definition of a particular concept, one will tend to notice the examples that fit it but be less attentive to those that don't fit. Similarly, at the level of theories of concepts, adherents of the classical view are likely to seek out those examples that (perhaps) fit the classical structure to the neglect of those that do not – hence the ubiquity of the example of BACHELOR in philosophy papers as a natural concept with a putative classical definition. Such cognitive biases do not change the fact, however, that classical definitions are in fact vanishingly rare – all the more so for high-level concepts like COGNITION.

Proposing and critiquing precise definitions of cognition is an appealing topic for academic discussion and publication, since there will always be something to say: extant proposals will always be subject to counterexamples – encompassing exemplars that they shouldn't include and/or missing

[1] These general points also provide strong reason to doubt the attainability of satisfactory conceptual analyses in contemporary philosophy, a point developed in Suhler (in preparation).

ones that they should. Such proposals and critiques are not, however, likely to be steps toward a fully adequate, precise definition of COGNITION. As discussed, efforts at conceptual analysis are probing mental concepts that likely have a non-classical structure and therefore are not amenable to the sort of precise demarcation being sought. As long as faith persists in science's need for precise definition, critiques of particular definitions simply perpetuate this misguided process. Therefore we will focus on criticizing the underlying rationale at a process level.

The Value of Associations

A further legacy of classical view of reasoning is a heavy emphasis of the role of logic and deductive reasoning in thought. Recent work, especially from Bayesian perspectives, strongly contradicts such views, arguing that thought is more statistical than formal-logical (Chater, 2009). High-level concepts likely deviate most strikingly from classical structure because their role is to cope with a massive, and poorly understood, world by holding vast numbers of statistically related ideas together into relatively few clusters. Crucially, this associative role does not require – and indeed can be undermined by – precise, exhaustive specification of conceptual content.

It is well known that exposure to a particular word or concept brings semantically related concepts sometimes including opposing concepts, into working memory. Thus subsequent thoughts are more likely to involve these related concepts. This priming role is important because our conceptual workspace is limited (e.g. Baars, 1997) and so there are finite number of concepts that have a chance, at any one time, to fit together into an explanation that will in turn fit with reality, producing a "eureka moment". As such, explanations will only avail themselves to us if a good number of the needed components are "close to mind". Useful high-level concepts bring to mind other concepts, such that as a whole the contents of the workspace are set up to yield explanations efficiently. Some authors go so far as to propose that a word's meaning is its statistical relations with other words (Lund, & Burgess, 1997), but even authors such as Barsalou (2008), who see little role for abstract symbol manipulation in cognition, acknowledge this power of words.

To see why this is of critical importance for science, consider an example from clinical psychology: what is meant when one asks whether a disorder such as depression is "psychological" or "biological". We would submit that the questioner is basically wondering whether the explanation (or "the story") of the disorder would be most usefully constructed using concepts that come to mind when one says "biology" such as HORMONES, and NEUROTRANSMITTERS, or couched in terms of DESIRES, OBSESSIONS, and RELATIONSHIPS, that come to mind when one says "psychology".

In this example, the cognitive utility of the concepts of PSYCHOLOGICAL and BIOLOGICAL lies not in their being explanatory themselves; rather, their value lies in their ability to organize groups of associated terms/concepts that may eventually figure into an explanation. We contend that this associative role of high-level concepts is, in the vast majority of cases, more important than their ability to support deduction, both in scientific and everyday inquiry. That is, these overarching concepts – what we are calling "high-level" concepts for short – are powerful drivers of association, but few play a direct role in mechanistic explanations; for example we seldom make strong, specific empirical claims about particular phenomena simply because "they are psychological/biological". Though we do not deny that concepts whose form allows for broad deduction are useful, it is not the case, as mentioned, that natural concepts typically have such a structure, nor it is it the case that concepts must be forced into such a form in order to help us to generate explanations.

Statistical knowledge must be relied upon in the early stages of investigation – when our investigatory approaches are based on hunches, necessarily entered into with highly imperfect knowledge of eventual results. For instance, EMBODIED COGNITION seems, at this stage, to be a concept that mainly serves to associate a number of disparate approaches motivated by the idea that an organism's physical form and environment cannot be abstracted away usefully from questions about cognition (Ziemke, 2003); it is, in essence, the broad claim that satisfying explanations of core cognitive phenomena will nontrivially involve bodies and their environs.

If we use high-level concepts to help us home in on explanatorily valuable associates, it's tempting to suggest that we should strive for greater conceptual precision. However, rather than being dictated by an imposition of formal definitions, semantic associations are built up over time by repeated co-activation of concepts, as during reading, conversation, or inner speech. Reflecting this history-dependent nature of association, individual researchers will, and should, tend organize their own workspace in a way that is suited to their particular problems. With the many kinds of cognitive science being done, the associations of cognition are bound to be promiscuous. We would simply add that this is not quite so undesirable as is often assumed.

Still, it is true that we associate precise ideas with good science. The great exemplars of scientific concepts are those that emerged as precise categories

(ATOM, MOLECULE, ENERGY, etc.). However, these concepts are precise because it is the nature of their referents to lend themselves to precision – they are putative "natural kinds". These conceptual success stories, coming mainly from physics, were not arrived at by tirelessly interrogating and refining preexisting ancient folk-physical concepts. Rather, precision in physical concepts was achieved through tireless empirical inquiry, and the positing of theories to make sense of these empirical observations.

Historical Evidence from Physics

Like most special sciences, the origins of physics were in philosophy; more specifically, it evolved out of what was called *natural philosophy* from antiquity through much of the 19th century. Initially, rather than there being any field called physics, which precisely defined its explananda and proceeded to explain them, there were instead a number of interests in specific natural (as opposed to man-made) phenomena. It was gradually realized, however, that explanations of such phenomena as heat, magnetism, and light could be made in terms of similar processes and entities. In more recent times, physics has become defined as the study of matter and energy (and even this definition may become obsolete), entities that actually would be rather foreign to the ancients from whose work the modern field descends: the Greeks thought that water fire, air, and earth were separate elements, of which all things were admixtures, including minds.

What the history of physics suggests is that fields of inquiry may discover what their explananda "actually are" as they move forward (Einstein & Infeld, 1961). The term "physics", rather than timelessly referencing a clearly defined set of concepts, questions and tools all revealed through analysis of the folk concept PHYSICAL, carries constantly updated statistical information about what ideas, facts, and phenomena have, to date, been found to "go together". What the ancients more likely had in common with us were very basic immediate sensory experiences, such as the sensations of light, sound and heat, of pressure, and of movement. These sensory experiences were what originally demanded explanation; however, these same entities have not ended up defining a coherent science. Rather, their study produced a cluster of explanatory concepts that continued to co-occur over and over again such that when one turned to be an important part of a particular story, another almost always did as well.

At any point in time, then, an attempt at precisely defining the subject matter of the field of physics would have proved descriptive rather than prescriptive. Paradigmatic examples of precise concepts arose quite anew out of a process of discovery, and might even have been closed off from discovery if the content and nature of the field had been fixed more than two millennia ago in the name of trying to precisely define the concept PHYSICAL.

Prospects for a Precise Definition of COGNITION

The pursuit of the unknown by any research method must be based on a guess at its potential benefit. It is impossible to prove, for example, that experimental work on any given question will yield enlightening results, but we intuit that it will because of experimentation's past record of success. In contrast, attempts to find, via conceptual analysis, precise definitions of concepts like EMOTION, SENSE, and COGNITION have so far not proven very successful. This may be because these concepts do not admit of precise definitions, or it may simply be that we have not achieved insights necessary for precise definitions. If precise definitions are "out there" then the experiences of other sciences give us the best means of guessing at how best to find them.

Experience in physics (see above) shows that intuitively appealing and precise definitions, when discovered, are extremely useful. But it is also clear that these come after great *empirical* effort, and are unlikely to correspond neatly to the natural concepts that a science has at its outset. So it might be that cognitive scientific concepts such as INTENTION, while intuitively appealing and arguably as real, psychologically, as HEAT, will not be foundational to the field's mature theories (for related discussion, see, e.g., Churchland, 1981; Thagard, 1990).

It is instructive to relate the above ideas to a recent proposal that the "mark of the cognitive" consists of *actions that have reasons* (Adams & Garrison, 2013). Recent research shows that humans seem to have very stable tendencies to perceive minds (and reasons/intentions) without much prompting. Seeing an entity as having reasons requires that different neural networks are engaged, and we then think about that entity differently than when we do not see it as motivated by reasons (Epley & Waytz, 2009). Thus, while perceptions of reasons behind actions may not be as experientially primitive as experiences of (say) heat and pressure, they are quite hard to avoid.

Nevertheless, despite the importance of our perception of minds and intentions in motivating interest in sciences that might help explain them, it is very possible that intentionality will end up taking a position in mature cognitive science is more analogous to that of heat, rather than that of energy in modern physics. For instance, recent influential proposals posit that human cognition is centrally concerned with explaining primary sensory data and this drives our neural activity, so that perceived minds, like objects, must be seen as part of our attempt at predicting the world (Friston & Frith, 2015). If so, then pursuit of deep principles of prediction, rather than of intentions, may eventually define our field. However, much like in physics, such sweeping redefinition will come *after* theories have proven their ability to explain empirical facts.

Finally it is worth noting that some commentators have argued that, in general, biology and cognitive science tend to provide us with explanations not via general laws (Bechtel, 2008), but rather by positing specific mechanisms and their interactions. Given that the discovery of law-like relations has driven the emergence of our most precisely defined high-level concepts from physics and physical chemistry (e.g. ELEMENT), sciences that do not produce laws would seem especially unlikely candidates to produce precise high-level definitions.

What Role for Conceptual Analysis in Science?

The inadequacy of the classical view of concepts and the nature of progress in other sciences (esp. physics) provide reasons to doubt the first key assumption we identified, at the outset, as underlying interest in defining terms like COGNITION: that an adequate, precise definition can be achieved if we think and argue hard enough about it. The associative role of concepts, meanwhile, provides reason to doubt the second key assumption – that precisely defining cognition will be a boon to cognitive scientific inquiry – since association does not require precise, fixed definitions.

Before elaborating on these points and why they cast doubt on the value of much conceptual analysis in cognitive science, it's worth pausing to emphasize that clarity about the meaning of concepts does have its uses. In particular, concepts that describe observed phenomena need to be defined clearly enough to allow interpretation and synthesis of empirical results. Most cognitive scientists have had the experience of going to read up on a particular research topic only to find that the key concept around which that topic is organized – IMITATION, EMPATHY, etc. – is used in a wide variety of ways in different papers and by different research groups. As mentioned briefly above, this variety itself is not necessarily a bad thing, so long as the authors of a given study are clear about how they are using a given concept and how their usage relates to other common usages in the literature. But when the time comes to reconcile and synthesize these results – a function typically performed by review articles – a degree of conceptual clarity is a must. Few things more effectively undermine the utility of a review article than inattention to the different ways in which a key term/concept is used in the body of research supposedly being reviewed. When this occurs, the review becomes little more than a bibliography, with the responsibility of achieving a degree of reconciliation and synthesis having been abdicated.

If conceptual clarity is important in these cases, then why not in the case of COGNITION? The reason, as already mentioned, is that unlike concepts such as IMITATION, EMPATHY, and ENERGY, high-level concepts like COGNITION and PHYSICAL do not figure very directly into investigations of specific phenomena. As with the concept PHYSICAL in physics, experiments and theories in cognitive science do not examine and explain cognition *qua* cognition; rather, they examine and explain more specific phenomena and processes.

Consider, by way of elaboration, an analogy to biology. Biology is, literally, the study of life, but biological experiments and theories are almost entirely concerned with more specific questions that in some way or another connect to living things. A precise definition of the concept LIFE is not going to help molecular geneticists or evolutionary theorists do their jobs better, since the concepts and methods that they use in day-to-day inquiry within their subdisciplines are much more specific – and empirically grounded – than that (see, e.g., Crick, 1966). Suggestions of and debate over a precise definition of LIFE are mostly philosophical curiosities, orthogonal to the methods, theories, and concepts that actually animate the daily work of scientific research. It is unclear why the great progress that biological science has made without a precise definition of its eponymous concept should necessarily be denied to a cognitive science that lacks a precise definition of COGNITION.

As we have emphasized throughout, the value of overarching disciplinary concepts like COGNITION and LIFE is likely to lie in their ability to organize mental workspaces and suggest connections between various lines of inquiry. Given this, the lack of sharp, pre-defined boundaries on what counts as cognition may actually *help* produce occasions for associations to be made and new lines of investigation to be opened. The potential restrictiveness of precisely fixing a definition of COGNITION can be seen by applying lessons from the history of physics to the history of cognitive science. As with physics in ancient times, a definition of COGNITION fixed in the early days of cognitive science (say, the late 1960s) would likely have centered upon logical symbol manipulation that goes on "inside the head". Such a definition would have closed off even connectionism as relevant to cognition, not to mention frameworks of embodiment (Varela, Thompson & Rosch, 1991;

Barsalou 2008), extended cognition (Clark, 1997), and metaphor (Lakoff & Johnson, 2008). These new frameworks expand the number of perspectives associated with COGNITION, thereby allowing further connections to be drawn and further lines of experimental inquiry to be opened up. The results of such inquiries, then, in turn, continue to modify our concept of COGNITION, and the process repeats itself.

Our argument, then, is that the value of the concept of COGNITION does not lie in any fixed definition, for such a definition will inevitably be beholden to the state of knowledge in the field and to the empirical and theoretical fashions of the day. Rather, concepts like COGNITION evolve with the field(s) in which they are used, providing a basis for associations and connections that generate empirical inquiries, and with them further conceptual modification. If a precise definition of COGNITION is possible at all, it will be achieved through a long, messy process of grappling with empirical reality, not traditional methods of conceptual analysis.

Conclusion

Both the allure and dangers of precise conceptual definitions are likely to be especially acute when a scientific field is in its early stages of understanding – as cognitive science currently is. Clarity and precision are often frustratingly hard to come by, a problem compounded in the case of cognitive science by the sheer complexity of the phenomena under investigation. Precise definitions are appealing in no small part because they promise such clarity and precision. The danger, however, comes from the near certainty that any definitions we strongly commit to at such an early stage will misguide future inquiry by ensconcing in those definitions all the empirical and theoretical limitations of the time at which they are fixed. For instance, even if our investigations ultimately yield an picture of cognitive systems that is not meaningfully "embodied" or "extended", the fact that we are so caught up with these ideas – ideas which would have been anathema to most cognitive scientists even 40 years ago – is surely a sign that we still have much to learn about what cognition is.

References

Adams, F., & Garrison, R. (2013). The mark of the cognitive. *Minds and Machines*, *23*(3), 339-352.

Baars, B.J., (1997). *In the theater of consciousness: The workspace of the mind*. Oxford Univ. Press.

Barsalou, L. W., Santos, A., Simmons, W. K., & Wilson, C. D. (2008). Language and simulation in conceptual processing. *Symbols, embodiment, and meaning*, 245-283.

Bechtel, W. (2008). *Mental mechanisms: Philosophical perspectives on cognitive neuroscience*. Taylor & Francis.

Churchland, P.M. (1981). Eliminative Materialism and the Propositional Attitudes. *Journal of Philosophy*, 77(2), 67-90.

Clark, A. (1997). *Being there: Putting brain, body, and world together again*. MIT Press.

Crick, F. (1966). Of molecules and men. University of Washington Press.

Einstein, A., & Infeld, L. (1961). *The evolution of physics: the growth of ideas from early concepts to relativity and quanta*. CUP Archive.

Epley, N., & Waytz, A. (2009). Mind perception. *Handbook of social psychology*.

Friston, K., & Frith, C. (2015). A Duet for one. *Consciousness and cognition*.

Griffiths, P. E. (1997). *What emotions really are: The problem of psychological categories* (p. 114). Chicago: University of Chicago Press.

Keeley, B. L. (2002). Making sense of the senses. *The Journal of Philosophy*,*99*(1), 5-28.

Lakoff, G., & Johnson, M. (2008). *Metaphors we live by*. University of Chicago Press.

Lund, C. & Burgess, K. (1997). Modelling parsing constraints with high-dimensional context space. *Language and cognitive processes*, *12* (2-3), 177-210.

Mervis, C. B., & Rosch, E. (1981). Categorization of natural objects. *Annual review of psychology*, *32*(1), 89-115.

Mulligan, K., & Scherer, K. R. (2012). Toward a working definition of emotion.*Emotion Review*, *4*(4), 345-357.

Oaksford, M., & Chater, N. (2009). Précis of Bayesian rationality: The probabilistic approach to human reasoning. *Behavioral and Brain Sciences*,*32* (1), 69-84.

Rosch, E., Thompson, E., & Varela, F. J. (1992). *The embodied mind: Cognitive science and human experience*. MIT Press.

Smith, E. E., & Medin, D. L. (1981). *Categories and concepts* (p. 89). Cambridge, MA: Harvard University Press.

Smith, R. (Ed.). (1997). *Aristotle. Topics. Books I and VIII: With Excerpts from Related Texts*. Oxford Univ. Press.

Thagard, P. (1990). Concepts and conceptual Change. *Synthese*, *82*(2), 255-274.

Rosch, Eleanor, Evan Thompson, and Francisco J. Varela. *The embodied mind: Cognitive science and human experience*. MIT press, 1992.

Ziemke, T. (2003, July). What's that thing called embodiment. In *Proceedings of the 25th Annual meeting of the Cognitive Science Society* (pp. 1305-1310). Mahwah, NJ: Lawrence Erlbaum.