Concepts in Change

Anna-Mari Rusanen, Samuli Pöyhönen*

Abstract

In this article we focus on the concept of concept in conceptual change. We argue that (1) theories of higher learning must often employ two different notions of concept that should not be conflated: psychological and scientific concepts. The usages for these two notions are partly distinct and thus straightforward identification between them is unwarranted. Hence, the strong analogy between scientific theory change and individual learning should be approached with caution. In addition, we argue that (2) research in psychology and cognitive science provides a promising theoretical basis for developing explanatory mechanistic models of conceptual change. Moreover, we argue that (3) arguments against deeper integration between the fields of psychology and conceptual change are not convincing, and that recent theoretical developments in the cognitive sciences might prove indispensable in filling in the details in mechanisms of conceptual change.

1 Introduction

For a philosopher of science interested in the dynamics and problems of interdisciplinarity, conceptual change research is a wonderful source of case studies. This is due to the interesting theoretical background of the field. Conceptual change research was originally a combination of three scientific fields developmental cognitive psychology, science education, and history and philosophy of science. This background has been a fertile foundation for a great number of studies on science learning, focusing on various different knowledge domains, at all educational levels from elementary school through college graduates.

These studies indicate that a crucial aspect of learning scientific topics involves radically altering the learner's prior conceptions in addition to cumulatively adding new knowledge to what is already there. This kind of learning process is what cognitive scientists and developmental and educational psychologists refer by the term ''conceptual change.''[1] When conceptual change occurs, a student does not merely accumulate more knowledge, but her conceptions of phenomena in a certain domain undergo a restructuring process that affects ontological commitments, inferential relations, and standards of explanation.

However, despite the abundance of studies on conceptual change a striking cluster of open questions remains. First, there is no agreement on what kinds of changes in belief and concept systems actually constitute conceptual change. Second, there is no consensus on what the specific mechanisms of conceptual change are, and thirdly, there is no common understanding of how to describe and model these mechanisms. A fundamental problem behind these questions is that there is no agreement on what it actually is that changes in conceptual change. As diSessa and Sherin (1998) put it, the trivial answer is ''concepts.'' However, this answer leaves us with the problem of saying what concepts are. In fact, as diSessa and Sherin observe, much of conceptual change research proceeds without a well-

* Both authors contributed equally to this work.

[1] See Strike and Posner (1982), Carey (1985), Posner et al. (1982), Thagard (1990), Chi (1992), Chi et al. (1994), Vosniadou and Brewer (1992)

defined notion of concept.

In this paper we aim to shed light on this central but elusive notion. Before we proceed, a couple of preliminary distinctions and clarifications are in order. In what follows, we focus on those models of conceptual change, in which conceptual change is understood as a form of learning process in an individual, not as scientific theory change. Moreover, our emphasis is not on theories of early learning, but on more advanced learning situations in the context of science education. In Sect. 2 we argue that theories of higher learning must often employ two different notions of concept that should not be conflated: psychological and scientific concepts. The usages of these two notions are partly distinct and thus straightforward identification between them is unwarranted. Hence, the strong analogy between scientific theory change and individual learning should be approached with caution. In Sect. 3 we argue that research in psychology and cognitive science provides a far more promising theoretical basis for developing explanatory mechanistic models of conceptual change. We aim to show that arguments against deeper integration between the fields of psychology and conceptual change are not convincing, and that recent theoretical developments in the cognitive sciences might prove indispensable in filling in the details of the mechanisms of conceptual change.

2 The Background: Two Notions of ''Concept''

2.1 Three Fields of Conceptual Change

Let us start with a brief historical overview. Conceptual change research can be seen to lie at the intersection of three scientific fields: science education, developmental cognitive psychology, and history and philosophy of science. Historically, an important factor behind the emergence of the field was the flood of research on misconceptions in the learning sciences in the 1980s (diSessa 2006; Vosniadou 1999). A crucial insight behind this movement was the idea that what makes domains such as energy, force, rational number, and evolution difficult to learn are the intuitive commonsense concepts that students have prior to instruction.[2] In various studies it was found that students often assimilate the scientific concepts presented to them in the classroom into their existing concepts. This often results in a web of ''misconceptions.'' Researchers in science education begun to identify and explore the structure, impact and organization of these misconceptions, and they started to develop instructional strategies that would succeed in helping students transform their intuitive concepts into more scientific ones. This research pointed to the existence of a specific kind of challenging learning task, one that required not only cumulatively increasing or revising one's theoretical beliefs, but also fundamental changes in the organization of those beliefs. This sort of learning was labeled ''conceptual change.''

2.2 Kuhn's Influence

The theoretical interpretations of these empirical findings were partially inspired by the history and philosophy of science. Arguably the most influential work there was Thomas Kuhn's historical analysis of the dynamics of scientific progress. Kuhn's *Structure of Scientific Revolutions* (1962) is emblematic of the approaches questioning the cumulative nature of theoretical and conceptual development of science. As is well known, according to Kuhn scientific revolutions follow a pattern. A dominant scientific paradigm—a basic way of conceptualizing, thinking, valuing, and emphasizing theoretically significant ideas—falls into a state of crisis, because it fails to provide explanations or solutions to significant new problems identified by the scientists. If an alternative paradigm with the potential to solve these problems is available, this increases the probability of a ''paradigm shift,'' i.e. universal adoption of the new framework. An important aspect of this picture of theory change was its emphasis on the difficulty of interpretation between paradigms: different paradigms often conceptualize their domains in

[2] See Strike and Posner (1982), Posner et al. (1982), Chi (1992), Vosniadou and Brewer (1992).

fundamentally different ways, and thus neat mappings between the concepts in different paradigms are hard to achieve.

Kuhn's model of scientific revolutions was an important source of theoretical ideas for the early conceptual change researchers. For example, George Posner and his colleagues—who originally introduced the term ''conceptual change'' in 1982—conceived learning process as being analogical to scientific conceptual change in the strictly Kuhnian sense (Strike and Posner 1982; Posner et al. 1982). They proposed that if a learner can solve problems within existing concepts and conceptions, she does not feel the need to change her current conceptions. Even when the conception does not successfully solve some problems of a new situation, the learner may often make only moderate changes to his beliefs and conceptions. According to Strike and Posner (1982), initiating a conceptually revolutionary process resulting in the replacement of the old paradigm with a new one requires a cognitive conflict, a profound dissatisfaction with the existing concepts. Naturally, the process requires also an alternative conception that is intelligible, appears initially plausible, and is believed to be a fruitful way to conceptualize the domain.

Even though many researchers have now rejected the details of Posner's and his colleagues' description of the process of conceptual change, the larger picture of the web of misconceptions as being analogical to paradigms in science is still quite widely accepted. It is a often assumed that the main reason for why students are so resistant to new scientific concepts is that their commonsensical beliefs and misconceptions organize and constrain learning in a manner similar to old paradigms in science, and that the process of conceptual change is as painful a process for a learner than a paradigm shift is for a scientific community in Kuhn's theory.

Kuhn's ideas also had repercussions in the third field of conceptual change, developmental cognitive psychology. In her early work, Susan Carey (1985, 1991) applied Kuhnian ideas of paradigm shifts and incommensurability in the context of children's conceptual change. More generally, in the 1980s, the field witnessed the rise of the analogy between early learning and scientific theory change in the form of *theory theory*. Theory theorists suggested that children's concepts are embedded within theories of different domains (e.g., knowledge about numbers, biology, physical world etc.), and that these theories develop (at least partly) independently from each other (cf. Keil 1989; Gelman and Coley 1991).

Perhaps the most radical form of theory theory advocates a strong analogy between learning and progress in science. According to the *continuity hypothesis* the basic processes of individual learning are similar to the processes of scientific development. An extreme example of this idea can be found in Alison Gopnik and Andrew Meltzoff's book *Words, Thoughts and Theories* (1997). According to the authors, learning in both science and human infants is based on the same cognitive processes (Gopnik and Meltzoff 1997, 7). In Gopnik and Meltzoff's view the most pivotal instrument for learning is theory change; both science and infants have the ability to replace one theory by another, if there is overwhelming evidence against the older theory and the evidence supports the new one. Moreover, Gopnik and Meltzoff claim that the processes underlying theory change are designed by the evolution, and their basic function is to allow humans to ''arrive at veridical conceptions of the world'' (Gopnik and Meltzhoff 1997, 15).

According to Gopnik and Meltzoff, when a child learns, she typically runs through a five-stage sequence of theory change. The child collects counter-evidence to the current theory, often denies the relevance of that counter-evidence, proposes auxiliary ad hoc hypotheses, formulates a new alternative theory, and tests the alternatives (Gopnik and Meltzoff 1997, 39–41). Moreover, the same is claimed to be true of scientists: also individual scientists go through all five stages in each episode of theory change. However, several authors have recently questioned this strong analogy between early learning and scientific theory formation. Fuller (2011) points out that whereas early learning is usually characterized as a strongly domain specific process, important episodes in scientific theory change are often results of ''cross-domain analogical inference'', of interdisciplinary transfer. This counts as evidence for the suspicion that there might be profound differences between conceptual change processes between the two domains. Moreover, the five-stage process overstates the role individuals play in scientific conceptual change. As Bishop and Downes (2002) point out, an individual scientist rarely passes through all of the stages in an episode of theory change. Instead, the history of science suggests that while some scientists collect counter-evidence to a dominant theory, others adamantly deny the existence of this evidence or its relevance to the topic. Moreover, whereas some scientists turn to formulating new theories, others typically propose ad hoc hypotheses to support the dominant theory. As Bishop and Downes suggest, these diverse reactions are a symptom of *the division of cognitive labor in science*. Science is not merely aggregated individual cognition and research is not sustained only by cognitive mechanisms inside scientists' heads (Nazer et al. 2002; Downes 1999; Bishop and Downes 2002). Instead, scientific research is essentially distributed cognition governed—and partly constituted—by social norms and mechanisms of social transmission. Moreover, as has been suggested in the research on the division of cognitive labor, the epistemic progress of a scientific community is often not maximized in a community of identical, perfectly rational individuals. Cognitive diversity, selective evidence uptake, and even irrational biases of individual scientists are often important preconditions for novel scientific discoveries (Kitcher 1993; Zollman 2007; Weisberg and Muldoon 2008).

2.3 Differences Between Psychological and Scientific Concepts

Based on the disanalogies between individual and scientific cognition discussed above, we suggest that scientific concepts should be clearly distinguished from children's or individual scientists' concepts. Also the literature on the history and philosophy of science supports this distinction: As the reflections on the division of cognitive labor suggest, scientific concepts are often treated primarily as constituents of scientific theories, and to serve this role, they should be thought of as objective and publicly shareable schemata, not essentially as mental representations of any particular individual.[3] This perspective suggests that scientific concepts could be understood as communally shared epistemic tools that scientists use to coordinate their efforts in their common task of knowledge production.

Clear examples of the distributed nature of scientific concepts can be found in the natural sciences, where research is often conducted by research groups rather than by individual scientists. There might be theoretical fields in physics, for instance, that no individual researcher grasps entirely. Instead, the knowledge concerning the phenomena can be distributed between different modeling perspectives and members of the research community. In such cases, the complete set of inferences licensed by the use of a concept in different contexts might not be completely grasped by any individual scientist but only by the research group taken as a whole. Furthermore, the objectivity and correctness of scientific inference are guaranteed by communication and error correction within the research group and within the wider scientific community. Importantly, this picture of scientific concepts applies also in less strongly distributed cases: what is referred to by speaking of scientific concepts are not mental representations of individuals but pieces of scientific knowledge that can be shared by a community of individuals.

2.4 Two Perspectives on Concepts

Edouard Machery (2009, ch. 2) has recently suggested a useful distinction that illuminates the differences between concept-talk in psychological and philosophical contexts. According to Machery, in the psychological sciences concepts are characterized as cognitive structures that are stored in long-term memory and that are used in the processes underlying higher cognitive abilities. Often these cognitive structures also play explanatory roles in psychological theories. Therefore, from this *mechanistic perspective*, concepts can be seen as structures that are used in causal explanations of individual behavior. In contrast, the *semantic perspective* on concepts usually employed in the

[3] See also Lappi (2011) for a similar analysis.

philosophical literature approaches concepts as constituents of propositional contents. This usage is avowedly non-mechanistic: a semantic theory of concepts cannot hope to do more than to correctly describe the conditions under which we would say that a person correctly uses a concept (cf. Machery 2009, ch. 2).

Machery's work suggests that the semantic and mechanistic perspectives on concepts are parts of two distinct epistemic projects, one explanatory and the other explicatory: As empirical sciences, cognitive psychology and cognitive sciences aim to explain the nature of cognitive systems and the structures and processes underlying them. The concept of ''concept'' plays a fundamental role in this endeavor. Briefly, in cognitive sciences concepts underlie complex cognitive states, such as thoughts or conceptions. Consequently, they are crucial to such psychological processes as categorization, inference, memory, learning, and decision-making, and often the precise meaning of ''concept'' is dependent on the particular experimental design at stake. The semantic theories, on the other hand, only attempt to explicate the conditions that would need to hold, were we to correctly describe someone as employing a certain concept x instead of a different concept y.

In our opinion the fundamental difference between these two perspectives suggests that the relationship between psychological and scientific concepts should be treated with caution. While there is no reason why scientific concepts could not be approached from the mechanistic perspective (and they have been, see Carruthers et al. 2002), we believe that the continuity hypothesis and other similar analogies might rely partly on conceptual confusion stemming from the conflation between these two perspectives. A straightforward identification between scientific concepts and psychological concepts is unwarranted and glosses over important conceptual distinctions. However, the differences between these two usages *do not suggest that only one notion of concept would be relevant* for the purposes of conceptual change research. In contrast, theories of conceptual change in science learning often need to employ both notions. Assessing whether a learner uses a concept correctly appears to be a task that must be approached from the semantic perspective: to use a scientific concept correctly is to know what inferences it participates in and it implies that one can coordinate one's own inferences with the collectively shared inferences, with scientific knowledge. Therefore, to determine whether someone has correctly learned a scientific concept, the inferences made by the person must be compared to the benchmark inferences implied by the normative scientific concept in question. On the other hand, in order to develop an explanatory theory of *how* a learner might achieve such coordination, we need to employ the mechanistic perspective. Learning is a psychological phenomenon and needs to be explained at the level of psychological and cognitive processes. We hold that a sophisticated theory of conceptual change should be committed to explicit mechanistic models of the theoretical constructs employed in research and test them against data. Thus, in addition to ''semantic'' descriptions of conceptual change at the level of (propositional) contents, genuinely explanatory theories should engage in characterization of conceptual change at the level of causal functioning of cognitive systems (cf. diSessa 2006; Clement 2008). In the next section we argue that the contemporary disciplines of psychology and cognitive science are perhaps the most promising contenders for sources of mechanistic evidence for theories of conceptual change.

3 Conceptual Change and Cognitive Mechanisms

If one seeks a characterization of concept that is appropriate for the explanatory purposes of learning research, an obvious option would be to look for it in the research on cognitive and developmental psychology of concepts. However, the fact that experimental psychology and conceptual change research have clearly distinct research interests and epistemic aims complicates this picture. Especially in the early experimental psychology research on concepts, attention was paid only to concepts that refer to classes of everyday entities such as birds and dogs. Concepts were often treated as constructs characterized only based on the task for which an experimental setting was designed for.[4] For example, if experimentalists were interested in categorization, i.e. determining whether an object belongs to one class or another, they defined concepts as categories. On the other hand, if they were interested in reasoning, they defined concepts as abilities to make inferences. As diSessa and Sherin (1998) remark, often nothing was said about how the task itself was accomplished. Typically it was only assumed that accomplishing certain experimental tasks is based on unified mental structures, and these mental structures were labeled as ''concepts.'' In other words, the concept of concept was often left as something of a black box (diSessa and Sherin 1998). According to what could be called the *vagueness argument*, the early psychological research on concepts was too incomplete to give an informative account of what kinds of entities concepts are. For this reason, it was fair to question to what extent this sort of experimental research on concepts offered knowledge that could be of use for the purposes of conceptual change research. In the end of this section, we suggest that contemporary research on conceptuality in psychology and cognitive sciences offers more rigorous, and thus more useful, information about the nature of concepts. However, we first deal with a more serious counter-argument against the relevance of psychology for conceptual change, *the grain size argument* (cf. Chi 2008).

3.1 The Grain Size Argument

The grain size argument is based on the observation that the central theoretical entities of psychological research and conceptual change reside on different levels of cognitive architecture. In conceptual change, especially in science learning, the focus is not on simple everyday concepts, but on complex, multifaceted and often highly abstract notions, such as the ''law of gravity'', ''force'' or ''computation.'' Moreover, the research on conceptual change often focuses on learning abstract and theoretical *knowledge domains* studied at schools or at universities. For this reason, in many studies the term ''conceptual change'' does not actually refer to change at the level of individual concepts at all (cf. Vosniadou and Brewer 1992; Chi 1992; Chi et al. 1994). Instead, it refers to the change, development, and reconstruction of systems of *conceptions*, which can be, for instance, schematically organized mental models (Vosniadou and Brewer 1992), categorical shifts (Chi 1992, 2008; Chi et al. 1994) or whole systems of beliefs. Because of this, the unit of conceptual change is often claimed to be at a larger grain size than the level of individual concepts (cf. Chi 2008).[5]

Our reply to the grain size argument comes in two parts. Firstly, there are accounts of conceptual change that locate conceptual change at the level of individual concepts, not in larger structures. For example, Gentner et al. (1997) state explicitly that there are three different grades of change: belief revision, theory change, and conceptual change. Conceptual change is ''a change in the fundamental concepts that compose the belief structure.'' Another example is early Carey (1985), who describes conceptual change as ''strong restructuring'', involving ''change at the level of individual concepts at the

---

[4] Since the rejection of the classical view of concepts in the 1970s, there have been three main approaches to concepts; the prototype view, the exemplar view, and the theory theory view of concepts. According to the prototype theory, concepts have a probabilistic structure. A concept of a class of objects is a prototype, a representation that contains statistical information about the properties possessed by (most of) the members of the class. On the contrary, the exemplar view considers concepts as representations of specific members of a category. These exemplars are thought to stand for the whole category. According to the theory–theory view, mental representations are similar to scientific theories and mental cognitive processes are analogical to scientific reasoning. Concepts are considered as embedded in theories about certain domains, as bodies of knowledge that contain causal, nomological, and functional generalizations about the corresponding categories (cf. Murphy 2004).

[5] Depending on the theory, the units of conceptual change differ. For example, in diSessas account the units of conceptual change are the coordination classes that organize information at the sub-conceptual level (diSessa and Sherin 1998, diSessa 1988, 1993). In Frank Keil's (1994) theory a similar role is played by modes of construal, and Stella Vosniadou (1992) employs the term 'framework.'

core of successive systems'' (p. 7). Thus, the grain size argument relies on a premise that is not unanimously shared among conceptual change researchers.

Our second reply is a bit more subtle, and its exposition spans over most of the remaining sections of our paper. In brief, the core idea is that even if theories of conceptual change were thought to focus on a larger and more complex grain size than the one of individual concepts, most of them are still at least implicitly committed to presuppositions concerning the nature of concepts. Typically these presuppositions consist of assumptions about the theoretical attributes of concepts (for example, whether concepts are discrete or continuous), and often they lie hidden in the details of theories of conceptual change. Often these presuppositions may seem to be quite innocent, but usually one should at least be aware of them. As we argue in the next section, theories of conceptual change may be implicitly committed to such views of cognition and cognitive operations that in turn may lead to very limited and strong views of the precise nature of the underlying cognitive mechanisms and cognitive architecture. The following example from Michelene Chi's early account of conceptual change (Chi 1992; Chi et al. 1994) is an illustration of possible interdependencies between theories of conceptual change and lower-level theories of cognitive architecture.

3.2 Conceptual Change as Ontological Shifts

The basic idea of Michele Chi's theory can be summarized as follows: Entities in the world belong to different ontological categories, such as MATTER, PROCESSES, and MENTAL STATES, and the learning process is a matter of organizing objects under the appropriate ontological categories. According to this view of conceptual change, the outcome of the learning process is a ''psychological tree of concepts'' in the learner's mind. For learning to be successful, this psychological tree should be *isomorphic* to the ''physical tree'' of categories in nature.[6] The basic learning mechanism underlying the transformation from the initial psychological tree to the outcome tree is a categorization process, i.e. a process of identifying or assigning a concept to a conceptual category to which it belongs (Chi 1992, 2008; Chi et al. 1994).

In this account, a concept—once categorized under superordinate concepts—can inherit features and attributions from its category membership, and thus the properties of objects falling under the concept can be learned by using background knowledge about the category. According to Chi, in addition to inheritance of properties from superordinate categories, the categorization model accounts for two further features of concept learning. First, when learners have no obvious basic category to assign to a new concept or a new phenomenon, they will assign it to the next appropriate higher-level category. Chi (2008) uses the following simple example: Suppose an observer in a museum sees a strange large creature (a gavial) with four short legs, scaly skin, and a flat bill-like snout. Because the observer does not know that the strange creature is a reptile, he/she will categorize it at the next level up, as a kind of animal, because it appears to have the properties of an animal. Second, when a learner relocates a concept from its initial tree to another one, an ontological shift happens. In Chi's model, this latter form of categorization leads to genuine conceptual change.

Although categorization plays a central role in Chi's theory, she does not say much about its precise nature. However, while she tries to remain noncommittal about the nature of processes underlying categorization, her learning model appears to presuppose some properties of category structure

[6] Curiously, a very similar view of nature was characteristic of classic Aristotelian and scholastic science. It was thought that all entities in reality had a place in a grand hierarchy of things called the tree of porphyry, and that it was the aim of science to uncover this true order of nature. In ninetieth century post-Darwinian biology, and consequently in other fields of science, the rise of population thinking lead to permanent abandonment of this essentialist view of nature, as it turned out that phenomena in nature could not be organized into such a hierarchical system (Hacking 2006). This suggests that Chi's early hierarchical ontologies cannot present a generally accurate view of the functioning of scientific concepts.

(categories are treated as property lists) (Chi 2008, 62). The argument by Eric Dietrich and Arthur Markman (2003) below suggests that even prima facie innocuous commitments about the nature of categorization might have far-reaching implications: depending on the notion of ''categorization'' one has in mind, one may end up with not only different theories of concepts, but also with completely different possible models of cognitive architecture. For example, if—as ultimately appears to be the case—Chi's account of conceptual change is dependent on the similarity-comparison picture of categorization, then it must be taken into account that it may be also committed to the view of concepts as discrete mental representations. Namely, Dietrich and Markman (2003) argue that a system is able to categorize environmental inputs only if it has internal states that impose classes of sameness on those inputs. Dietrich and Markman's reasoning goes as follows: Categorization requires that inputs are identified and classified. Hence, discriminating inputs is necessary for categorizing and it requires enduring classes of sameness (achieved by, for instance, chunking). This sort of chunking, in turn, seems to require discrete representations, i.e representations that are bounded and uniquely identifiable. Therefore, if a system can categorize on the basis of similarity comparisons, then it operates with discrete representations.

Suppose, at least for the sake of argument, that Dietrich and Markman's reasoning is sound. It suggests that employing Chi's categorization model of learning has implications on one's possible views of cognitive architecture. Commitment to discrete representations rules out non-representationalist theories of cognitive architecture such as dynamical systems theory, which is based on continuous representations (Van Gelder and Port 1995). It also perhaps rules out some connectionist models that operate with highly distributed representations. In a similar way, other models of conceptual change will have their own consequences. For instance, according to the theory theory views, such as Susan Carey's account (1985, 2000, 2009), categorization is not based on similarity comparisons, but on reasoning processes that are analogical to inference to the best explanation (Murphy and Medin 1985) and other forms of explanatory, inductive or probabilistic reasoning (Smith et al. 1992). Such an account differs radically from the similarity-comparison based account described above, and may be committed to a completely different picture of the human cognitive architecture.

As such, these constraints on cognitive architecture do not present a problem for theories of conceptual change. However, the examples given here aim to illustrate that often even minor theoretical decisions can have important interdisciplinary ramifications: When developing a model of conceptual change, once its structure and basic learning mechanisms are specified, the model will be compatible only with some theories of cognitive architecture, not all of them. This conclusion appears to reflect a general property of interdisciplinary knowledge production: Scientific disciplines are rarely completely insulated from each other, but instead evidence and theories accumulated in a particular research program usually constrain the space of theoretical possibilities in other scientific fields. The grain-size argument appears to stem from insufficient attention to the inter-dependencies between phenomena occurring at different levels of cognitive systems.

3.3 Missing Mechanisms

Chi's account of categorization is also more generally a nice illustration of the need for an interdisciplinary approach to conceptual change. Chi describes categorization just by saying that ''[c]ategorizing is the process of identifying or assigning a concept to category to which it belongs'' (Chi 2008, 62). She does not appear to offer any description of how ''identifying'' or ''assigning'' actually happens, or what kind of cognitive mechanisms they actually are. Based on a large contemporary literature in the philosophy of science, we contend that if there is no specification of these tasks at the level of cognitive operations responsible for these cognitive phenomena, these terms offer quite shallow understanding of the underlying processes and do not qualify as descriptions of cognitive mechanisms. We hold that to be able to really *explain* conceptual change, one should give an account of how certain cognitive mechanisms sustain or produce the transformation from the initial state, the common sense picture of the world, to the outcome state, internalization of scientific concepts. Genuinely explanatory models are ones in which the phenomenon is explained by giving an accurate and sufficient description of how a causal mechanism, a hierarchical system composed of component parts and their properties, sustains or produces the phenomenon (Bechtel and Richardson 1993; Machamer et al. 2000; Craver 2006, 2007). In addition, genuine explanations offer the ability to say not merely how the system in fact behaves, but to say how it would behave under a variety of circumstances or interventions (Craver 2006, 2007; Woodward 2003).

Also in the literature on conceptual change it is common to make reference to mechanisms of conceptual change: Categorization (Chi), differentiation (Carey), coalescence (Carey), branch jumping (Thagard), and tree switching (Thagard) are often described as central learning mechanisms responsible for conceptual change. The problem is that often these purported ''mechanisms'' of conceptual change are not sufficiently specified, but they are mere preliminary sketches of possible mechanisms. In other words, even if the operations above are often referred to as mechanisms, they often fail to satisfy the requirements for genuine mechanism descriptions: A causal mechanism is a complex system that produces certain behavior because of the causal interactions between its components. In mechanism descriptions these components and their interactions must be characterized spatially, temporally and causally accurately and in sufficient detail (Machamer et al. 2000; Craver 2006, 2007). If these aspects are not specified, the purported ''mechanisms'' include what could be called *filler terms*. Filler terms describe the relation between the input and the output of the process, but they offer little specific information of how the change was brought about. If a model includes filler terms and is thus an incomplete model of mechanisms, it should rather be called a ''mechanism sketch'' (Craver 2006, 2007).

In the sciences studying complex systems one is often stuck with only partial information concerning the targets of research, and thus with mechanism sketches. However, the explanatory usefulness of mechanism description decreases, when the incompleteness of a model increases. Filler terms are often barriers to scientific progress when they veil failures of understanding (Craver 2006, 2007). If, for example, the term ''assign'' is used to stand for a process with largely unknown properties and implementation, we don't really explain what happens, but we only have an illusion of explanation (Craver 2006; Rozenblit and Keil 2002). In sum, having numerous loose filler terms in an explanation often threatens to undermine its explanatory power. In the following penultimate section of our paper we suggest that contemporary psychology and cognitive science offer promising resources for filling in the filler terms in theories of conceptual change.

3.4 Towards an Interdisciplinary Account of Conceptual Change

As Clement (2008) remarks, there are very few, if any, models of conceptual change, in which the mechanisms of conceptual change are specified in sufficient detail.[7] This suggests that many of the current accounts might not in a strict sense qualify as substantial causal theories sufficient for explanation and manipulation of learning phenomena involving conceptual change. Instead, building adequate models might require attention to the implementing cognitive processes.[8] Thus, even if the units of conceptual change were thought to lie at a larger grain size than the level of individual concepts, and for that reason psychology and cognitive sciences might prima facie appear irrelevant, these disciplines may turn out to be of great importance, because the details of mechanisms of

[7] Susan Carey's (2009) work is one of the few examples of conceptual change research that explicitly aims to uncover the psychological mechanisms underlying conceptual change.

[8] It is not completely clear, how to define the notion of ''cognitive mechanism'' (see Lappi and Rusanen 2011). Lappi and Rusanen suggest that in some cases explanatory models in cognitive science may contain non-implemented, abstract mechanisms.

conceptual change may have to be specified at the level of cognitive processes studied by these disciplines. This concludes our second reply to the grain size argument: we suggest that understanding the evidential and explanatory interdependencies between the different sciences related to conceptual change could be cashed out as a project of weaving an inter-level mechanistic picture of human conceptual abilities (cf. Craver 2007).

At least conceptual change research would greatly profit from a genuinely interdisciplinary approach. There is plenty of research done in cognitive sciences that could offer specification for terms such as ''categorization'', ''identifying'' and ''assigning'', and conceptual change research might benefit from closer cooperation with these fields of research. For instance, categorization mechanisms have been studied and modeled both in deductive and statistical manner (Goodman et al. 2008). A recent development in the statistical tradition is rational analysis of categorization in the context of concept learning (Tenenbaum and Griffiths 2001). These analyses show how several important aspects of categorization and concept learning can be described in terms of Bayesian inference, and how simple concepts can be learned by rational inductive learning, if there is a given and fixed hypothesis space. Moreover, in cognitive science there is an interesting line of research, in which scientists try to model and explain conceptual development in Bayesian terms. For instance, Alison Gopnik and Joshua Tenenbaum with their colleagues offer examples of research on cognitive mechanisms that in an interesting way integrate problems of conceptual development to the grand theoretical picture of the cognitive sciences (Gopnik 2011; Sobel et al. 2004).

Generally, one of the most challenging problems for the whole field of statistical learning is to integrate the models with the more complex cases. The current statistical learning models of concepts are better suited for describing simpler observational concepts. In general, these current models work well for simple everyday concepts like birds or dogs. However, even if some recent accounts of statistical learning models try to capture features of learning complex concepts (Tenenbaum and Griffiths 2001; Griffiths et al. 2008; Goodman et al. 2008), it is still unclear how to extend those accounts to cover learning very complex and highly abstract concepts such as the concept of ''computation'' or the concept of ''gravity'' (Goodman et al. 2008). For this reason, there is still plenty of work to be done, if one aims to extend these statistical models of learning to cover relevant cases of conceptual change, in which the focus is on complex concepts with dynamic, multidimensional and rich constituent structures. For example, physics concepts are often multi-faceted in the sense that they cannot be associated with any particular set of properties, and it is unclear what kind of composition they actually have (Koponen and Pehkonen 2010).

If the attempts to integrate the statistical models with the general picture of conceptual change turned out to be unsuccessful, this would suggest that some of our learning mechanisms can be described by using Bayesian models while some others cannot. Actually, this is what the current empirical evidence on concept learning in psychology seems to suggest. Contemporary psychological research suggests that our capacity for conceptual thought appears to require a wide range of skills such as fast categorization, inductive inference, concept acquisition, and compositionality (Smith et al. 1992; Barsalou 1990; Osherson et al. 1990). Perhaps this heterogeneous set of capacities is sustained by many cognitive mechanisms and all of them are needed to explain our conceptual abilities (Machery 2009).

For research on conceptual change, this means that *different instances of cognitive change in early childhood, science learning, and scientific cognition could be sustained by entirely different mechanisms*. If this is the case, a cluster of new research questions arises. If there are several learning mechanisms, we should aim to describe which mechanisms are used in which particular learning situation, and explain why a particular mechanism is activated in certain situations. In addition, we should be able to answer the question of whether the different learning mechanisms are domain-specific, and whether they are architecturally ''innate'' features of human cognition. Moreover, if learning requires the orchestrated functioning of a group of cognitive mechanisms, we should be able to explain how these learning mechanisms interact.

We envision that this kind of knowledge on the specific causal mechanisms underlying learning would be important for the future of conceptual change. If we were able to individuate and identify the causally relevant mechanisms underlying complex, higher-level learning, this would give us more satisfactory theoretical knowledge of the studied phenomena and it would allow us to develop our methods of instruction in an efficient manner.

4 Discussion

There is an enormous amount of studies of students' ability to learn scientific content. These studies carried out in a multiplicity of different contexts consistently shows that a crucial aspect of learning many elementary topics in science involves radically altering the learner's prior conceptions in addition to adding new knowledge to what is already there. However, even within the literature devoted on conceptual change, there is no agreement on what actually changes during episodes of conceptual change and how the change happens. Instead, the field consists of a kaleidoscopic variety of different theoretical perspectives (diSessa 2006). We suggest that when developing genuinely explanatory models of conceptual change, it is often necessary to describe the phenomena at the level of underlying cognitive mechanisms. For this reason, future conceptual change research may have to rely more on the theoretical and methodological work done in cognitive sciences and experimental psychology.

Acknowledgments

We thank Otto Lappi, Petri Ylikoski and the anonymous referees for useful comments on an earlier version of this paper.

References

Barsalou, L. W. (1990). On the indistinguishability of exemplar memory and abstraction in category representation. In T. K. Srull & R. S. Wyer (Eds.), *Advances in social cognition, volume III: Content and process specificity in the effects of prior experiences* (pp. 61–88). Hillsdale, NJ: Lawrence Erlbaum Associates.

Bechtel, W., & Richardson, R. (1993). *Discovering complexity, decomposition and localization as strategies in scientific research.* New Jersey: Princeton University Press.

Bishop, M., & Downes, S. (2002). The theory theory thrice over: The child as scientist, Superscientist or social institution? *Studies in the History and Philosophy of Science*, 2002(33), 121–136.

Carey, S. (1985*). Conceptual change in childhood*. Cambridge, MA: MIT Press.

Carey, S. (1991). Knowledge acquisition: Enrichment or conceptual change? In S. Carey & R. Gelman (Eds.), *The epigenesis of mind* (pp. 257–291). Hillsdale, NJ: Lawrence Erlbaum Associates.

Carey, S. (2000). Science education as conceptual change. *Journal of Applied Developmental Psychology*, 21(1), 13–19.

Carey, S. (2009). *The origin of concepts*. Oxford: Oxford University Press.

Carruthers, P., Stich, S., & Siegal, M. (Eds.). (2002). *The cognitive basis of science*. Cambridge: Cambridge University Press.

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

Chi, M. T. H. (2008). Three types of conceptual change: Belief revision, mental model transformation, and categorical shift. In S. Vosniadou (Ed.), *Handbook of research on conceptual change* (pp. 61–82). Hillsdale, NJ: Erlbaum.

Chi, M. T. H., Slotta, J. D., & de Leeuw, N. (1994). From things to processes: A theory of conceptual change for learning science concepts. *Learning and Instruction*, 4, 27–43.

Clement, J. (2008). The role of explanatory models in teaching for conceptual change. In S. Vosniadou (Ed.), *International handbook of research in conceptual change* (pp. 417–452). New York: Routledge.

Craver, C. F. (2006). When mechanistic models explain*. Synthese*, 153, 355–376.

Craver, C. F. (2007). *Explaining the brain: What a science of the mind-brain could be.* New York: Oxford University Press.

Dietrich, E., & Markman, A. B. (2003). Discrete thoughts: Why cognition must use discrete representations. *Mind and Language, 18*(1), 95–119.

diSessa, A. A. (1988). Knowledge in pieces. In G. Forman & P. Pufall (Eds.), *Constructivism in the computer age* (pp. 49–70). Hillsdale, NJ: Erlbaum.

diSessa, A. A. (1993). Toward an epistemology of physics. Cognition and Instruction, 10(2, 3), 105–225. diSessa, A. A. (2006). *A history of conceptual change research: Threads and fault lines*. In K. Sawyer (Ed.), Cambridge handbook of the learning sciences (pp. 265–281). Cambridge, UK: Cambridge University Press.

diSessa, A. A., & Sherin, B. (1998). What changes in conceptual change? *International Journal of Science Education*, 20(10), 1155–1191.

Downes, S. (1999). Can scientific development and children's cognitive development be the same process? *Philosophy of Science*, 66, 565–578.

Fuller, T. (2011). Is scientific theory change similar to early cognitive development? Gopnik on science and childhood. *Philosophical Psychology*, forthcoming.

Gelman, S. A., & Coley, J. D. (1991). Language and categorization: The acquisition of natural kind terms. In S. A. Gelman & J. P. Byrnes (Eds.), *Perspectives on language and thought: Interrelations in development* (pp. 146–196). Cambridge: Cambridge University Press.

Gentner, D., Brem, S., Ferguson, R. W., Wolff, P., Markman, A. B., & Forbus, K. D. (1997). Analogy and creativity in the works of Johannes Kepler. In T. B. Ward, S. M. Smith, & J. Vaid (Eds.), *Creative thought: An investigation of conceptual structures and processes* (pp. 403–459). Washington, DC: American Psychological Association.

Goodman, N. D., Tenenbaum, J. B., Feldman, J., & Griffiths, T. L. (2008). A rational analysis of rule-based concept learning. *Cognitive Science*, 32(1), 108–154.

Gopnik, A. (2011). Probabilistic models as theories of children's minds. *Behavioral and Brain Sciences,* 34(4), 200–201.

Gopnik, A., & Meltzoff, A. (1997). *Words, thoughts and theories*. Cambridge: MIT Press.

Griffiths, T. L., Christian, B. R., & Kalish, M. L. (2008). Using category structures to test iterated learning as a method for identifying inductive biases. *Cognitive Science*, 32(1), 68–107.

Hacking, I. (2006). Des classifications naturelles (pp. 422–430). lectures at College de France. http://www.college-de-france.fr/media/historique/UPL17441_ihackingcours0506.pdf.

Keil, F. (1989*). Concepts, kinds, and cognitive development*. Cambridge, London: MIT Press.

Keil, F. C. (1994). The birth and nurturance of concepts by domains: The origins of concepts of living things. In L. A. Hirschfeld & S. A. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture.* New York: Cambridge University Press.

Kitcher, P. (1993). *The advancement of science. Science without legend, objectivity without illusions*. Oxford: Oxford University Press.

Koponen, I., & Pehkonen, M. (2010). Coherent knowledge structures of physics represented as concept networks in teacher education. *Science & Education*, 19, 259–282.

Kuhn, T. (1962). *The structure of scientific revolutions*. Chicago: University of Chicago Press.

Lappi, O. (2011). How not to study conceptual change. In F. Seroglou, V. Koulountzos & A. Siatras (Eds.) *The Proceedings of 11th International IHPST and 6th Greek History, Philosophy of Science Teaching Joint Conference*, Epikentro Publishers.

Lappi, O., & Rusanen, A.-M. (2011). Turing machines and causal mechanisms in cognitive science. In P. McKay Illari, F. Russo, & J. Williamson (Eds.), *Causality in the sciences* (pp. 224–239). Oxford: Oxford University Press.

Machamer, P., Darden, L., & Craver, C. F. (2000). Thinking about mechanisms. *Philosophy of Science,* 67(1), 1–25.

Machery, E. (2009). *Doing without concepts.* Oxford: Oxford University Press. Murphy, G. L. (2004). The big book of concepts. Cambridge: MIT Press.

Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review,* 92(3), 289–316.

Nazer, D., Ruby, A., Nichols, S., Weinberg, J., Stich, S., Faucher, L., et al. (2002). The baby in the lab-coat: Why child development is not an adequate model for understanding the development of science. In P. Carruthers, S. Stich, & M. Seagal (Eds.), *The cognitive basis of science* (pp. 335–362). Cambridge: Cambridge University Press.

Osherson, D., Smith, E. E., Wilkie, O., & Lopez, A. (1990). Category-based induction. *Psychological Review,* 97(2), 185–200.

Posner, G. J., Strike, K. A., Hewson, P. W., & Gertzog, W. A. (1982). Accommodation of a scientific conception: Towards a theory of conceptual change. *Science Education*, 66(2), 211–227.

Rozenblit, L., & Keil, F. (2002). The misunderstood limits of folk science: An illusion of explanatory depth. *Cognitive Science*, 26, 521–562.

Smith, E., Langston, C., & Nisbett, R. E. (1992). The case for rules of reasoning. *Cognitive Science*, 16, 1–40.

Sobel, D., Tenenbaum, J., & Gopnik, A. (2004). Children's causal inferences from indirect evidence: Backwards blocking and Bayesian reasoning in preschoolers. *Cognitive Science*, 28(3), 303–333.

Strike, K. A., & Posner, G. J. (1982). Conceptual change and science teaching. *European Journal of Science Education*, 4(3), 231–240.

Tenenbaum, J., & Griffiths, T. (2001). Generalization, similarity, and Bayesian inference. *Behavioral and Brain Sciences,* 24, 629–641.

Thagard, P. (1990). Concepts and conceptual change. *Synthese,* 82, 255–274.

van Gelder, T., & Port, R. (1995). It's about time: Overview of the dynamical approach to cognition. In R. Port & T. van Gelder (Eds.), *Mind as motion: Explorations in the dynamics of cognition* (pp. 1–43). Cambridge: Bradford Books/MIT Press.

Vosniadou, S. (1992). Knowledge acquisition and conceptual change. *Applied Psychology: An*

International Review, 41(4), 347–357.

Vosniadou, S. (1999). Conceptual change research: State of the art and future directions. In W. Schnotz, S. Vosniadou, & M. Carretero (Eds.), *New perspectives on conceptual change* (pp. 3–13). Amsterdam: Elsevier Science.

Vosniadou, S., & Brewer, W. F. (1992). Mental models of the earth: A study of conceptual change in childhood. *Cognitive Psychology*, 24, 535–585.

Weisberg, M., & Muldoon, R. (2008). Epistemic landscapes and the division of cognitive labor. *Philosophy of Science*, 76, 225–252.

Woodward, J. (2003). *Making things happen. A theory of causal explanation*. Oxford University Press. Zollman, K. (2007). The communication structure of epistemic communities. Philosophy of Science, 74(5), 574–587.

.