Toward a Peircean Theory of Human Learning: Revealing the Misconception of *Belief Revision*

Mary Keeler and Uta Priss

Ostfalia University of Applied Sciences
Wolfenbüttel, Germany
mkeeler@uw.edu
www.upriss.org.uk

Abstract. *Belief Revision* was conceived to model how humans *do* think, and has found application in machine learning. This paper argues that Peirce's theory of inquiry conceives how we *must* think, if we want to keep improving our knowledge. Distinguishing between these two views, psychological (empirical) and pragmatic (normative), is crucial to our improvement of human learning methodology, especially as we develop interactive engagement methods for learning STEM concepts. Examining efforts to model *Belief Revision* in AI can reveal the limitations of this conceptualization for human learning, due to its misconception of Peirce's pragmatic theory of inquiry.

1 Introduction

In AI research, "Belief Revision" refers to the process of changing beliefs to accommodate new information, as part of the knowledge representation challenge to model learning by machine cognition. This research originates in early pragmatist philosophy, especially C.S. Peirce's essay, "The Fixation of Belief" [1877] and Dewey's learning theory. AI researchers have adopted Peirce's term "abductive reasoning" to explain belief revision as "guessing right" or "inference to the best explanation." Peirce considered abduction (or retroduction) to be a valid form of logical reasoning, but logical empiricists in cognitive science consider it to belong properly to psychology. Although he struggled to distinguish his logic of reasoning (as normative science) from psychology (as empirical science), the significance of this distinction remains unexamined, leaving cognitive research fundamentally confused about belief and reasoning in the process of human learning.

If, as Peirce explains, "the essence of belief is the establishment of a habit, and different beliefs are distinguished by the different modes of action to which they give rise" ["How to Make Our Ideas Clear" (1878)], and yet the purpose of logical reasoning (as inquiry, or learning) is to find the truth (which is an ideal limit we never reach, but must hope will keep our reasoning effectively progressing), then belief and reasoning are at odds in purpose. In fact, we might say that belief must be suspended during logical reasoning, so that it does not "block the way of inquiry," in Peirce's terms. Perhaps inquiry suspends belief, and belief suspends inquiry? If so, we might clarify the confusion in cognitive research by distinguishing belief from reasoning to enable more effective study of human learning. This effort begins to clear the way toward improvement upon the prevailing constructivist theory of learning, especially

H.D. Pfeiffer et al. (Eds.): ICCS 2013, LNAI 7735, pp. 193–209, 2013.

for teaching STEM subjects, by replacing Dewey's version of pragmatism with Peirce's pragmatic theory of inquiry—to advance the *how* toward the *why* of learning. The following sections respond to basic questions: What is the difference between belief and reasoning? What role can belief have in learning? Can we learn to use beliefs effectively in learning?

2 Belief Revision Theory: From Philosophy to AI

Tracing the history of *Belief Revision (*BR) development reveals the theoretical misunderstanding among philosophers, cognitive and computer scientists, linguists, and economists that we should expect in interdisciplinary research. Many in AI who have developed learning systems that incorporate some version of BR have ignored conceptual difficulties identified by philosophers. Even researchers who explicitly address conceptual problems, such as P. Thagard, have increased the confusion by simplifying theoretical fundamentals in their early work. The complexity of issues in that history is beyond this paper's scope, but we briefly cover its origins and evolution.

Although *Frontiers in Belief Revision* [Williams and Rott 2001] informs us that BR theory first came into focus in the work of the philosophers W. Harper [1976; 1977] and I. Levi [1977; 1980; 1991], we find evidence in [Doyle] that it began to take shape in the work of H. Kyburg [1961] who, in his delineation of the fundamentals from antiquity, erroneously conflates Peirce's with Dewey's pragmatism.

> There are two fundamentally distinct ways of thinking about thinking about the world. One is contemplative; the other is oriented toward action. One seeks pure knowledge; the other is pragmatic. One leads to hedged claims; the other leads to categorical claims in a hedged way. Both approaches to thinking about the world have ancient roots: Socrates, seeking wisdom; Alexander, the man of action. Both are represented in contemporary philosophy: Carnap wanted to associate with each statement of our language its appropriate degree of confirmation, relative to what we know; Peirce and Dewey took the impetus for deliberation to be the necessity to choose an action, and the outcome to be the act. Both approaches are represented in artificial intelligence: the probabilists taking the correct representation of our trans-evidential conclusions about the world to be hedged statements (the probability of rain tomorrow is .67), and the logicists taking the representation to be categorical statements (it will rain tomorrow), appropriately hedged in a non-monotonic logic—the conclusions can be withdrawn in the face of new evidence. [Kyburg 1994: 1-2]

In 1985, C. Alchourrón, P. Gärdenfors, and D. Makinson introduced their model (AGM) based on their earlier work (1978-82) [see Gärdenfors 1992, 2011], to address a critical problem. In simple terms, their work responds to: "how do you update a database of knowledge in the light of new information? What if the new information is in conflict with something that was previously held to be true? An intelligent system should be able to accommodate all such cases" (using an established set of postulates). The AGM model represents beliefs as sentences in some formal language that does not capture all aspects of belief:

> The beliefs held by an agent are represented by a set of such belief-representing sentences. It is usually assumed that this set is closed under logical consequence, i.e., every sentence that

> follows logically from this set is already in the set. This is clearly an unrealistic idealization, since it means that the agent is taken to be "logically omniscient." However, it is a useful idealization since it simplifies the logical treatment; indeed, it seems difficult to obtain an interesting formal treatment without it. In logic, logically closed sets are called "theories". In formal epistemology they are also called "corpora", "knowledge sets", or (more commonly) "belief sets." [*Stanford Encyclopedia of Philosophy* (*SEP*): plato.stanford.edu/]

Levi [1984, 1986] clarified the nature of this idealization, pragmatically (in Dewey's sense of "pragmatic"): "a belief set consists of the sentences that someone is *committed to believe*, not those that she actually believes in" [*SEP*]. According to Levi's analysis, we are logically committed to believe in all the consequences of our beliefs, but typically our performance does not live up to this logical commitment. The belief set (as the set of an agent's epistemic commitments) is therefore larger than the set of her actually held beliefs. C. Misak's review of Levi's analysis hints at the confusion in its philosophical derivation:

> Levi's approach to the revision of belief and the growth of knowledge belongs, as he says, in the pragmatist tradition. Such an approach to epistemology emphasizes the context of inquiry and is teleological and decision oriented. Revisions of knowledge—scientific or otherwise—are taken to be central, and they are value laden in the sense that they are always made relative to the aims the agent is committed to promoting and to the agent's existing corpus of belief. Levi's decision theory is erected on these pragmatist foundations, and it promises to go a long way in clarifying how we should conduct our inquiries. But with respect to the foundations, the relationship between Levi and his predecessor Peirce is one of coincidence rather than supersession. [Misak, 264]

The original core assumption of belief revision was *minimal change*: the knowledge before and after the change should be as similar as possible. In the case of update, this principle formalizes the assumption of *inertia*. In the case of revision, this principle enforces the assumption of preservation of much information as possible in the change. In 1990, Thagard surveyed and assessed accounts of the psychological functions of concepts, and suggested that "conceptual change can come in varying degrees, with the most extreme consisting of fundamental conceptual reorganizations. … Understanding epistemic change requires appreciation of the complex ways in which concepts are structured and organized and of how this organization can affect belief revision" [255]. Researchers in AI began describing their work in terms of combining BR and *abduction* [see Santos 1991; Boutilier and Becher 1993; survey of methods in Walliser, Zwirn, and Zwirn 2004]. Thagard took his work beyond belief revision, in *Conceptual Revolutions* [1992], to the question of revolutionary conceptual change.

> Theorists in philosophy, psychology, and artificial intelligence have proposed different views of the nature of concepts. A rich account of concepts and conceptual change is needed to overcome the widely held view that the growth of scientific knowledge can be understood purely in terms of belief revision with no reference to conceptual change. Concepts serve many psychological functions, and can be understood as complex computational structures organized into kind-hierarchies and part-hierarchies. Such structures involve rules that can combine to explanations. [33]

Based on his earlier work, *Computational Philosophy of Science* [1988], Thagard advocated that techniques derived from AI could be used "to understand the structure and growth of scientific knowledge," and reciprocally, "The theory of revolutionary conceptual change developed is germane to central issues in cognitive psychology and artificial intelligence, as well as to the philosophy of science" [3]. Thagard explained what had become the accepted view of *abduction* in AI—at the core of confusion between psychological and logical views of inference (emphasis added to points that will be examined, below).

> The problem of inference to explanatory hypotheses has a long history in philosophy and a much shorter one in psychology and artificial intelligence. Scientists and philosophers have long considered the evaluation of theories on the basis of their explanatory power. In the late nineteenth century*, C.S. Peirce discussed two forms of inference to explanatory hypotheses:* hypothesis*, which involved the acceptance of hypotheses, and* abduction*, which involved merely the initial formation of hypotheses* (Peirce 1931-1958; Thagard 1988). Researchers in artificial intelligence and some philosophers have used the term "abduction" to refer to both the formation and the evaluation of hypotheses. [62: Thagard lists Pople 1977, Peng and Reggia 1990, Josephson et al. 1987, and Hobbs, Stickel, Appelt 1990, and Martin, Charniak and McDermott 1973, 1986]

Thagard proposed a theory of explanatory coherence (TEC) as "central to the general theory of conceptual change in science," which would "account for a wide range of explanatory inferences," in terms of principles to encompass the considerations "that suffice to make the judgments of explanatory coherence." He demonstrated the sufficiency of these principles by implementing his theory in a connectionist computer program called ECHO, which was applied to "complex cases of scientific and legal reasoning" [62-63; and see Thagard 1978: "Best Explanation: Criteria for Theory Choice"; and Van Fraassen's "The Pragmatics of Explanation"].

We abbreviate Thagard's summary of BR history in the following outline comparing philosophical to cognitive science perspectives and warning of terminological confusion.

1. Contemporary analytic philosophers take sentences to be the objects of epistemological investigation.
2. Knowledge is something like true justified belief, so increasing knowledge entails adding to what is believed.
3. Epistemology primarily evaluates strategies for improving stocks of beliefs, construed as sentences or attitudes toward sentence-like propositions (e.g., Gärdenfors models an individual's epistemic state as a consistent set of sentences that can change by expansion and contraction) [1992, 19].
4. Cognitive psychologists pay less attention to BR and far more attention to "what is the nature of concepts?"
5. Cognitive researchers in AI often follow philosophers' analysis of BR, but also pay attention to how knowledge can be organized in conceptual structures, or *frames* (Minsky 1975; for reviews see Thagard 1984, 1988).
6. Even philosophers who take cognitive science seriously consider BR to be the center of epistemology and pay little attention to conceptual change (e.g., A. Goldman 1986). The central question for epistemology has been: "when are we justified in adding and deleting beliefs from the set of beliefs judged to be known?" Epistemology should also address another question: "what are concepts and how do they change?" Concepts are relevant to epistemology if the question of conceptual change is not identical to the question of belief revision. [Based on Thagard 1992, 20]

Thagard's thesis of non-identity between conceptual change and belief revision did not attract serious response from belief-revision theorists, which W. Park [2010] finds "especially curious in view of the fact that theory of belief revision—with the AGM paradigm at its core—has over the past two decades expanded its scope far beyond epistemic logic and philosophy of science to include computer science, artificial intelligence, and economics." Various critiques of BR have appeared [see Aliseda 1997; Boutilier, Friedman, and Halpern 1998, 2008; Nebel 1989: Rott 2000, Darwich and Pearl 1997; Friedman and Halpern 2000; Gilles 2002; Nayak et al. 2003; van Benthem 2004; Jin and Thielscher 2007; Olsson and Enqvist 2011]. In their more recent critique, Friedman and Halpern [2008] identified methodological problems, and argued that careful study of belief change will require explicit ontology or scenario representation of the process.

By the late 1990s, it became clear that BR theory could be related to formal learning theory, as K. Kelly explains in his rationale for bridging between the two theories in "The Learning Power of Belief Revision" [1998, 111]:

> The guiding principle of belief revision theory is to change one's prior beliefs as little as possible in order to maintain consistency with the new information. Learning theory focuses, instead, on learning power: the ability to arrive at true beliefs in a wide range of possible environments. … learning power depends sharply on details of the methods. Hence, learning power can provide a well-motivated constraint on the design and implementation of concrete belief revision methods.

Perhaps, BR research has been a constructive model of itself? Certainly it has raised core questions that reveal its own theoretical confusion: What is the motivation for revising beliefs (*why* change), beyond maintaining consistency and coherence, and what constitutes new information (*why* is it selected)? Attempts to model theory change provoked the need to explain the *how* of explanatory *coherence*, exposing deeper questions of *why*, what motivates consistency and coherence in representation? While AI research has focused on the *how* questions, the *why* questions have been neglected, and must be addressed in learning theory.

3 Belief as an Instinct

The evolutionary study of human cognition is also an interdisciplinary challenge (involving cognitive and computer sciences, philosophy, economics, and linguistic anthropology). Chomsky's and Pinker's theories of "the language instinct" are now well known [Pinker and Bloom 1999], and can be traced back to W. von Humboldt in the eighteenth century, whose ideas were embraced by nineteenth century anthropologists [Humbolt 1999].

The earliest book-length account of "Evolutionary Psychology" as a discipline is an undergraduate text by evolutionary epistemologist H. Plotkin. His *Evolution in Mind: an Introduction to Evolutionary Psychology* [1998] traces its origins back to Darwin and contemporaries, carried forward by the early pragmatists such as James. Plotkin explains that the early use of *instinct* to account for human behavior was an irresponsible extension of Darwin's theory of the continuity among species.

> The hunt for human instincts around the turn of the [18th] century and during its first and second decades marks a low point in human sciences. Without empirical or theoretical justification of any kind, thousands of human instincts were invented, many of them extraordinarily trivial and silly. Worse still, some writers attributed putative characteristics of whole nations to instincts. … ideology and, in this case chauvinism, intruded into the application of a concept derived from evolutionary theory to human psychology. This early phase of ascribing human action to instincts cannot be called either science or psychology of any description. Although William James himself had come, after a time and with some qualms, to champion the idea of the existence of at least some human instincts, even his great reputation could not save so weak a conceptual edifice. The net effect of the work of the eugenicists and instinct theorists during this sorry episode in the human sciences was not only to discredit the idea of instincts, but by association, seriously to weaken the influence of evolutionary ideas within psychology. [28-29]

Workman and Reader further explain:

> … the concept of instinct was dropped from social scientists' terminology in the twentieth century partly because it was considered too imprecise a term to be scientifically meaningful (see Bateson, 2000). Furthermore, many so-called instinctive behaviours are capable of being modified by experience, in which case it is difficult to see where instincts finish and learning begins. A final reason why the concept of instinct fell out of favour is that a new approach to the social sciences denied their existence and saw culture rather than biology as being the principal determiner of human behaviour [11-12].

J. Tooby and L. Cosmides [1992] took issue with the dominant, non-evolutionary model in the social sciences (Standard Social Science Model) for its assumptions, which in turn were a reaction to the preceding biological determinist assumptions [see Workman and Reader 2004, 12]. F. Coolidge and T. Wynn, in *The Rise of Homo Sapiens: The Evolution of Modern Thinking* [2009], explain the motivation for evolutionary psychologists, who over the last two decades "have used reverse engineering to argue for the selective reasons behind a large array of human cognitive abilities, including spatial cognition (Irwin Silverman and Eals, 1992), language (Steve Pinker, 1997), cheater detection (Leda Cosmides, 1989), and even religion (Pascal Boyer, 2001)." All are convinced that "the current structure of human cognition preserves traces of its evolutionary past," features of an "earlier evolutionary adaptedness." A major tenet of evolutionary psychology is that our minds are adapted to a time when humans lived in small hunting and gathering groups, not the modern world, which helps explain many current psychological problems.

Workman and Reader mention that the study of genetics was dominated by "DNA-thinking," but that recently many researchers have conceived non-DNA methods of heritability, in the new field of *epigenetics* ("so new there is still no generally accepted definition of it") [52]. They find support in epigenetics for their proposal: "We now hypothesize that some neural mutation or epigenetic event led to a reorganization of the brain that enabled modern thinking" [55].

Another recent critical introductory text [Swami 2011] evaluates research from "the last decade of dramatic change in our understanding of the way in which the mind operates and the reasons behind a myriad of human behaviours." Evolutionary psychological explanations have supplanted the traditional idea that "nurture trumps nature" in human behavior by positing that shared mental architectures govern our behavior.

> Evolutionary psychology (EP) tries to identify human psychological traits that are evolved adaptations … Applying the same adaptationist thinking about physiological mechanisms common in evolutionary biology, evolutionary psychology argues that the mind has a modular structure similar to the body's. Different modular adaptations serve different functions, so that much of human behavior is the evolutionary result of psychological adaptations to solve recurrent problems in human ancestral environments. As an effort to integrate psychology into the other natural sciences, EP understands psychology as a branch of biology. … a framework that not only incorporates the evolutionary sciences on a full and equal basis, but that systematically works out all of the revisions in existing belief and research practice that such a synthesis requires. [11]

Most recently, this research trend toward "behavioral genetics" has encouraged some to conclude that beliefs are genetically determined in brain function, not directly but through traits that are, like personality. Psychologist M. Shermer [2011] argues (in *The Believing Brain: how we construct beliefs and reinforce them as truths*) that we may like to think our beliefs come from experience, but instead they come first and then we devise reasons for believing. Our brains are "belief-generating machines," to avoid uncertainty and find patterns to follow. Even scientists operate under paradigms, but science has "built-in self-correcting mechanisms that check belief claims. … Most guesses are false-positive (low-cost errors), … even scientists start out with beliefs, which they then try to justify" [278].

Shermer claims that, without science ("the ultimate bias detection machine"), our brains convince us that we are always right. He describes a dozen major tendencies in judgment (biases and effects) identified by researchers.

The Confirmation Bias *(The Mother of All Cognitive Biases*, because it gives birth in one form or another to most of the other heuristics)**:** the tendency to seek and find confirmatory evidence in support of already existing beliefs and ignore or reinterpret disconfirming evidence.

Hindsight Bias (a type of *time-reversal* confirmation bias): the tendency to reconstruct the past to fit with present knowledge. Once an event has occurred, we look back and reconstruct how it happened.

Self-Justification Bias (related to the hindsight bias): the tendency to rationalize decisions after the fact to convince ourselves that what we did was the best thing we could have done. Once we make a decision about something in our lives we carefully screen subsequent data and filter out all contradictory information related to that decision, leaving only evidence in support of the choice we made.

Attribution Bias (several kinds: situational, dispositional, intellectual, and emotional): the tendency to attribute different causes for our own beliefs and actions than that of others (common in political and religious beliefs).

Sunk-Cost Bias**:** the tendency to believe in something because of the cost sunk into that belief.

Status Quo Bias**:** the tendency to opt for whatever it is we are used to, that is, the status quo.

Endowment Effect**:** the tendency to value what we own more than what we do not own.

Framing Effects: the tendency to draw different conclusions based on how data are presented. Framing effects are especially noticeable in financial decisions and economic beliefs.

Anchoring Bias**:** the tendency to rely too heavily on a past reference or on one piece of information when making decisions, when we have no objective anchor for comparison.

Availability Heuristic**:** the tendency to assign probabilities of potential outcomes based on examples that are immediately available to us, especially those that are vivid, unusual, or emotionally charged, which are then generalized into conclusions upon which choices are based.

Representative Bias (related to the availability bias)**:** the tendency to judge an event probable to the extent that it represents the essential features of its parent population or generating process.

Inattentional Blindness Bias**:** the tendency to miss something obvious and general while attending to something special and specific. [Based on Shermer 259-272, and he lists 25 additional biases.]

If our beliefs cause instinctive behavior, evolved under conditions we no longer need to respond to, how can they be "updated," or what is their role in learning?

4 A Peircean Theory of Learning?

No doubt many belief biases and effects were operating among BR researchers, but even our superficial untangling of their confusion points to the pervasive influence of Levi's theory of inquiry (based on Dewey's pragmatism), as R. Hilpinen reminds us:

> Peirce's account of abduction and induction as the main forms of non-demonstrative reasoning has inspired Levi's theory of inquiry and belief revision, articulated in several recent publications [Levi 1997; 1991; 1996; 2000]. In contemporary methodology, abduction is generally recognized as a distinctive form of reasoning, and models of abductive reasoning are being studied in applied logic, cognitive science and artificial intelligence, and in the theory of diagnostic reasoning. (See Josephson and Josephson, 1994; Magnani et al., 1999; Gabbay et al., 2000; Flach and Kakas, 2000.) [2004, 652]

We identify three fundamental ways that Levi's influence prevents BR from effectively modeling human learning as *the improvement of knowledge*, rather than as merely updating a database of biased beliefs. *Learning as inquiry* must challenge assumptions, not "fix" them by maintaining their consistency and coherence as we experience new information. Levi's misconceptions can instruct us how to make better use of Peirce's theory of inquiry.

1. Levi confuses Peirce's *fallibilism* with Popper's *falsificationism* [see Levi 1984, 112], a pervasive problem in philosophy and consequently in AI, especially in efforts to model scientific inquiry. As explained in [Keeler 2008]: "Peirce's inductive fallibility is a metaphysical condition, not to be confused with Popper's falsification, which is strictly a deductive procedure (see Haack, *Evidence and Inquiry*, p. 131)." Misak charges, "Isaac Levi uses C. S. Peirce's fallibilism as a foil for his own 'epistemological infallibilism'" [256]. Levi insists that both the proximal and ultimate purpose of inquiry is to eliminate error (to produce true, maximally consistent belief systems [Levi 1991]). As Misak points out, Peirce's "critical commonsensism" agrees with Levi that we do not doubt what we believe; but she clarifies:

> by "infallibilism" [Peirce] means the position which is opposed to his own fallibilism; the position that our beliefs (or at least some of them) are incorrigible, or not the sort of things that are ever in need of revision. Fallibilism insists that an inquirer must "be at all times ready to dump his whole cartload of beliefs, the moment experience is set against them" (Peirce 1931, 1.55). He cannot have "any such immovable beliefs to which he regards himself as religiously bound to be loyal" (Peirce 1931, 6.3). Such an attitude would block the path of inquiry because our minds would be closed, and hence, we would never be motivated

enough to inquire. One of Peirce's reasons for endorsing fallibilism is the fact that our faculties sometimes fail us, and we cannot be sure when these failures occur. … Even the greatest mathematicians, he notes, are susceptible of making the simplest mistakes in arithmetic—all it takes is a little lapse of attention. [1987, 259]

2. Levi misconceives Peirce's *abduction* [even dismisses it, see Note 1], which results from his misinterpretation of Peirce's fallibilism and his reliance on Peirce's early essay on inquiry, "The Fixation of Belief" [see Kasser 2011]. Misak describes Peirce's early "doubt-belief" model of inquiry.

The notion of inquiry is central in Peirce's epistemology. He characterizes it as the struggle to rid ourselves of doubt and achieve a state of belief. An agent has a body of settled belief: a set of statements which are not, in fact, doubted. Statements in this body, however, are susceptible to doubt, if it is prompted by some "positive reason," such as a surprising or recalcitrant experience. A body of settled belief is presupposed for the operation of inquiry in that there has to be something settled for surprise to stir up. Doubt is not voluntary, and hence, we cannot simply do it, as Descartes suggests, at will. But when it impinges upon us, it "essentially involves a struggle to escape it" (Peirce 1931, 5.372, n.2) and so, as soon as we are thrown into doubt, inquiry is ignited. It continues until we reach a settled belief—a belief that we regard as "infallible, absolute truth." So Peirce characterizes the path of inquiry as follows: settled belief, doubt, inquiry, settled belief. [259]

However, Peirce's abduction evolved with his theory of inquiry [see Anderson 1986; and for discussions of Peirce's mature theory of abductive reasoning, see Hintikka, 1998, 2007, Hilpinen 2004, and Kapitan 1997]. After careful consideration of BR theory's interpretation of abduction, J. Hintikka concludes that abduction "cannot be thought of as an inference to the best explanation" [2007, 42]. A. Aliseda even advocates finding new terminology for abduction in AI [2010, 9]. Hintikka returns to Peirce's own notion of inference, for clarification: "I call all such inference by the peculiar name, abduction, because its legitimacy depends upon altogether different principles from those of other kinds of inference" [*Collected Papers (CP)* 6.524 (1901]. He points to T. Kapitan's summary of those "different principles."

(1) Inference is a conscious, voluntary act over which the reasoner exercises control (5.109, 2.144).
(2) The aim of inference is to discover (acquire, attain) new knowledge from a consideration of that which is already known (MS 628: 4).
(3) One who infers a conclusion *C* from a premise *P* accepts *C* as a result of both accepting *P* and approving a general method of reasoning according to which if any *P*-like proposition is true, so is the correlated *C*-like proposition (7.536, 2.444, 5.130, 2.773, 4.53–55, 7.459, L232:56).
(4) An inference can be either valid or invalid depending on whether it follows a method of reasoning it professes to and that method is conducive to satisfying the aim of reasoning—namely, the acquisition of truth (2.153, 2.780, 7.444, MS 692: 5). [Kapitan, 479; in Hintikka 2007, 44]

Hintikka explains that Peirce is "going beyond rules of inference that depend on the premise-conclusion relation alone and is considering also rules or principles of inference 'of an altogether different kind.' These rules or principles are justified by the fact that they exemplify a method that is conducive to the acquisition of new knowledge." Furthermore:

> the validity of an abductive inference is to be judged by *strategic* principles rather than by *definitory* (move-by-move) rules. This is what makes an abductive inference depend for its legitimacy "upon altogether different principles from those of other kinds of inference." What these "different principles" were in Peirce's mind can be gathered from his various statements. One typical expression of the difference is Peirce's distinction between the validity and the strength of an argument. … it is only in Deduction that there is no difference between a valid argument and a strong one ("Pragmatism as the Logic of Abduction," p. 17). Thus an argument can be logical but weak. [2007, 44-45]

Hintikka illustrates the "vantage point" of this "interrogative approach":

> Peirce's terminology can be claimed merely to follow ordinary usage when he calls an interrogatively interpreted abductive step an inference. The reasoning of the likes of Sherlock Holmes or Nero Wolfe is not deductive, nor does it conform to any known forms of "inductive inference." The "deductions" of great detectives are in fact best thought of as question–answer sequences interspersed with deductive inferences (I have argued). Yet people routinely call them "deductions" or "inferences" accomplished by means of "logic" and "analysis." They now turn out to be right strategically speaking, though not literally (definitorily) speaking. From the strategic vantage point, we can say thus that any seriously asked question involves a tacit conjecture or guess. [2007, 55]

Hilpinen agrees and further explains:

> Peirce's distinction between abduction and induction has sometimes been associated with the logical empiricist's distinction between the context of discovery (the discovery or invention of an explanatory hypothesis) and the context of justification (the confirmation or disconfirmation of a hypothesis by empirical evidence) [Reichenbach 1938]. Many logical empiricists regarded only the latter as a proper subject of logical and philosophical investigation, and thought that the study of the discovery of hypotheses belongs to psychology rather than logic. It is clear that Peirce's rules of abduction [**see Note 2**] can be said to "justify" a hypothesis in the way in which inductive reasoning can justify its conclusions: a good abduction justifies a hypothesis as a potential explanation worthy of further empirical testing. In Peirce's words, we can say that *abduction justifies an interrogative attitude towards a hypothesis*. … an abduction leads to a "conjecture" and can justify only an "interrogative" attitude towards a proposition. According to Peirce, "Induction shows that something actually is operative, Abduction merely suggests that something may be." [652; *CP* 5.171 (1903); emphasis added]

Kapitan's careful analysis concludes that Peirce's abduction as a form of valid inference forces us to broaden the concept of validity [2004, 491] in Peirce's theory of inquiry: "[abduction] is the only logical operation which introduces any new ideas; for induction does nothing but determine a value, and deduction merely evolves the necessary consequences of a pure hypothesis" [*CP* 5.171 (1903); and see "Grounds of Validity of the Laws of Logic …" *CP* 5.341-357 (1868-93)].

3. Levi's BR theory was originally conceived for and applied to changes in belief of a single individual and in a computerized database. S. Hansson explains five major differences in modeling of scientific knowledge processes that traditional BR theory fails to account for:

The Processes of Change are Collective
The Data/Theory Division
A Partly Accumulative Process

> *Explanation-Management Rather than Inconsistency-Management*
> *The Irrelevance of Contraction*
>
> The transformation of high probabilities to full belief can be described as a process of uncertainty-reduction, or "fixation of belief" (Peirce 1877). It helps us to achieve a cognitively manageable representation of the world, thus increasing our competence and efficiency as decision-makers. This transformation is just as necessary in the collective processes of science as it is in individual cognitive processes. In science as well, our cognitive limitations make it impossible to keep track of an extensive net of interconnected probabilities. We cannot (individually or collectively) deal with a large body of human beliefs such as the scientific corpus in the massively open-ended manner that an ideal Bayesian subject would be capable of. As one example of this, since all measurement practices are theory-laden, no reasonably simple account of measurement would be available to a Bayesian approach (McLaughlin 1970). [In Olsen and Enqvist 2011, *Belief Revision Meets Philosophy of Science,* 48-50.]

However, as Levi and Hintikka agree, "Epistemologists ought to care for the improvement of knowledge rather than its pedigree" [Levi 1980, 1; Hintikka 2007]. Encouraging that direction, T. Deacon's *Incomplete Nature: How Mind Evolved from Matter* gives us a neuroscientist's examination our current "ecology" of cognition.

> People tend to be masters of believing incompatible things and acting from mutually exclusive motivations and points of view. Human cognition is fragmented, our concepts are often vague and fuzzy, and our use of logical inference seldom extends beyond the steps necessary to serve an immediate need. This provides an ample mental ecology in which incompatible ideas, emotions, and reasons can long co-exist, each in its own relatively isolated niche. Such a mix of causal paradigms may be invoked in myths and fairy tales, but even here such an extreme discontinuity is seldom tolerated. Science and philosophy compulsively avoid such discontinuities. More precisely, there is an implicit injunction woven into the very fabric of these enterprises to discover and resolve explanatory incompatibilities wherever possible, and otherwise to mark them as unfinished business. Making do with placeholders creates uneasiness, however, and the longer this is necessary, the more urgent theoretical debate or scientific exploration is likely to be. [2011, 63]

Peirce even eliminates belief from the collaborative learning in science.

> Full belief is willingness to act upon the proposition in vital crises, opinion is willingness to act upon it in relatively insignificant affairs. But pure science has nothing at all to do with action. The propositions it accepts, it merely writes in the list of premisses it proposes to use. Nothing is vital for science; nothing can be. Its accepted propositions, therefore, are but opinions at most; and the whole list is provisional. The scientific man is not in the least wedded to his conclusions. He risks nothing upon them. He stands ready to abandon one or all as soon as experience opposes them. Some of them, I grant, he is in the habit of calling established truths; but that merely means propositions to which no competent man today demurs. It seems probable that any given proposition of that sort will remain for a long time upon the list of propositions to be admitted. Still, it may be refuted tomorrow; and if so, the scientific man will be glad to have got rid of an error. There is thus no proposition at all in science which answers to the conception of belief. [CP 5.635 (1898)]

Extending this view, we argue that Peirce's theory of scientific inquiry represents the logical (not psychological) essence of learning as *collectively engaging in the deliberate, continuous improvement of knowledge*. Hintikka's interpretation of Peirce's abduction clarifies it as a *strategic* procedure for gaining self-critical control of our belief biases, rather than as mere guessing. Furthermore, we must point out, Peirce's abduction also gives us the true "Mother of All Beliefs," the pragmatic aim (or *normative constraint*), the "why" that motivates all learning: the tendency to *hope* that we can *continue to improve* knowledge, by *engaging with the community of inquirers* [see *CP* 5.311 (1878)].

5 Didactic Implications

STEM subjects (science, technology, engineering and maths) are often difficult to teach as demonstrated, for example, by high drop-out and failure rates among first year university students. A significant amount of research has been dedicated to understanding why such subjects are difficult to teach and learn and how to help students in such subjects. As an example, Physics Education Research has come to the conclusion that students have pre-existing "misconceptions" about physics concepts that are counter-intuitive [Hestenes et al., 1992]. Students can usually learn to operate with such concepts in formulas (and thus pass exams) but if their understanding of such concepts is questioned, they fail [Hake, 1998]. Physicists have developed "concept inventories" [Hestenes et al., 1992] which are lists of questions about difficult concepts expressed mostly in everyday language. Using these concept inventories one can measure how much students know at the start and end of a semester. The learning gain in introductory physics courses measured in this manner is often very small if the courses employ standard teaching methods, but apparently students learn much more if "interactive engagement" teaching methods are used [Hake, 1998].

Interactive engagement methods include "peer instruction" (where students explain and discuss concepts with each other [Mazur, 1996], "problem-based" or "inquiry-based" learning, "flipped classroom" (where students read the lectures at home and practice during the lecture time) and "just-in-time teaching" (where lecturers respond to student questions instead of presenting a fixed lecture). But not all of these methods are guaranteed to be successful. For example, Loviscach [2012] reports that his flipped classroom with video-recorded lectures is very popular with students but has not led to significant improvements of students' marks. Noschese and Burk [Noschese, 2011] coin the term "pseudoteaching" for teaching that is on the surface very good, well liked by students and staff and where students think that they are learning a lot, but does still not lead to deep and substantial learning. Thus while not all interactive engagement teaching methods are successful, some are. Therefore the question arises as to what is the reason for the success of some methods.

It should be stressed that this paper is concerned with the learning and understanding of difficult concepts in STEM subjects. Other types of learning (such as learning a skill or learning vocabulary) might require different teaching approaches. But with respect to students overcoming their pre-existing misconceptions, the problem appears to be essentially the problem of "making ideas clear" and "fixing beliefs" as discussed by Peirce. Thus our hypothesis is that teaching methods are successful in changing students' beliefs if they encourage students to conduct inquiry in Peirce's sense. On the one hand, a better understanding of how to practically use Peirce's inquiry could make it easier for educators to predict which teaching methods are likely to be successful. On the other hand, educational methods that help students overcome misconceptions could be studied as examples of successful inquiry. At the moment the main philosophical grounding of interactive engagement methods appears to be constructivism [Ben-Ari, 1998], which is certainly an improvement over Cartesian views of science but could be further improved by a *Peircean* pragmaticist view.

6 Conclusions

According to Peirce, belief relies on assumptions—the attenuation of doubt; while reasoning progresses by suppositions—the perpetuation of doubt. A belief is a cognitive rule (or habit telling us how to think) for guiding action; while reasoning constructs rules of thought to relate facts (explaining why repeatable observations fit together). A belief can be accepted without regard for facts (without asking "why") and, to the extent that such habits are not consciously formed, they can become addictive. Recent evolutionary views of cognition have encouraged psychologists to consider whether our belief capability is instinctive, a form of adaptivity that often limits learning. While we can reason to construct rules to be tested in "learning by experience," belief can lead us to misjudge experience: making us overconfident about what we know, risk-averse to searching for disconfirming evidence, and prone to interpret evidence to preserve established beliefs. Peirce explains that in ordinary everyday life we act from instinct, with just enough reasoning necessary to connect these habits of thought with specific occasions. In learning, however: "when ones purpose lies in the line of novelty, invention, generalization, theory—in a word, improvement of the situation … —instinct and the rule of thumb manifestly cease to be applicable. The best plan, then, on the whole, is to base our conduct as much as possible on Instinct, but when we do reason to reason with severely scientific logic" [CP 2.176-78 (1902)]. That logic accounts for learning in the conduct of inquiry as the tasks of abduction, induction, and deduction.

We encourage Conceptual Structures researchers to pursue Peirce's theory of inquiry as a logical, rather than psychological, basis for pragmatically (strategically) improving our ability to learn. Hintikka's suggestion of "games of inquiry" [2007:183, 222], and Gärdenfors's "geometry of thought" [see 2000] inspire our hope that Conceptual Structures research can make significant contributions in the future [and see Keeler 2007, 2008, 2010]. Friedman and Halpern's [2008] call for "explicit ontology or scenario representation" in BR presents an invitation to research how a Peircean theory might lead us to more effective human learning.

Note 1. In *The Fixation of Belief and Its Undoing, Changing Beliefs Through Inquiry* [1991], Levi even dismisses abduction: "Peirce devoted substantial effort to characterizing the differences between deduction, abduction, and induction as differences in the formal structure of arguments. At the beginning of the twentieth century he abandoned the project. It would be useful if contemporary writers would take the lessons Peirce learned nearly a century ago to heart." [See p. 172; he gives no references for this claim, but should have been aware at least of Peirce's 1908 essay, "Neglected Argument for the Reality of God," which features his stages of inquiry (*CP* 6.468-473).]

Note 2. Hilpinen explains that Peirce's "logic of abduction," as rules for good abductions, was an important aspect of his pragmatism. "An abduction is an inference which leads to a conjectured explanation, thus the logic of abduction may be expected to include conditions of adequacy for explanatory hypotheses as well as rules of discovering explanatory hypotheses."

Rule of Abduction 1: The hypothesis (the "conclusion" of an abduction) must be capable of being subjected to empirical testing.

Rule of Abduction 2: The hypothesis must explain the surprising facts. Peirce observes that an explanation may be a deductive explanation which renders the facts "necessary," or it may make the facts "natural chance results, as the kinetic theory of gases does" [*CP* 7.220]. These rules have counterparts in the more recent theories of explanation, for example, in Carl G. Hempel and Paul Oppenheim's account. RA1 and RA2 correspond to Hempel and Oppenheim's "logical conditions of adequacy" for scientific explanations [Hempel and Oppenheim 1948/65, 247-248]. Peirce's logic of abduction also contains a rule which may be called the Principle of Economy:

Rule of Abduction 3: In view of the fact that the hypothesis is one of innumerable possibly false ones, in view, too, of the enormous expensiveness of experimentation in money, time, energy and thought, is the consideration of economy. Now economy, in general, depends upon three kinds of factors: cost, the value of the thing proposed, in itself, and its effect upon other projects. Under the head of cost, if a hypothesis can be put to the test of experiment with very little expense of any kind, that should be regarded as giving it precedence in the inductive procedure. [*CP* 7.220 n. 18); Hilpinen, 651-52]

References

Alchourrón, C., Gärdenfors, P., Makinson, D.: On the Logic of Theory Change: Partial Meet Contradiction and Revision Functions. Journal of Symbolic Logic 50(2), 510–530 (1985)

Aliseda, A.: Abduction Reasoning, Logical Investigations into Discovery and Explanation. Studies in Epistemology, Logic, Methodology, and the Philosophy of Science, Synthese Library (2010), http://www.scribd.com/doc/8145990/Aliseda-Abductive-Reasoning

Amini, M.: Logical Machines: Peirce on Psychologism. Disputatio: International Journal of Philosophy II, 335–348 (2008)

Anderson, D.: The Evolution of Peirce's Concept of Abduction. Transactions of the Charles S. Peirce Society 22(2), 145–164 (1986)

Ben-Ari, M.: Constructivism in Computer Science Education. SIGCSE Bulletin 30(1), 257–261 (1998)

van Benthem, J.: Dynamic Logic for Belief Revision. Journal of Applied Non-Classical Logics 17(2), 129–155 (2004)

Bacchus, F., Grove, A., Halpern, J., Koller, D.: A Response to Believing on the basis of evidence. Computational Intelligence (1994) (Article first published online April 2, 2007), doi:10.1111/j.1467-8640.1994.tb00141.x

Boutilier, C., Becher, V.: Abduction As Belief Revision: A Model of Preferred Explanations. In: AAAI 1993 Proceedings, pp. 642–648 (1993)

Coolidge, F., Wynn, T.: The Rise of Homo Sapiens: The Evolution of Modern Thinking. Wiley-Blackwell (2009)

Cosmides, L., Tooby, J.: The Adapted Mind: Evolutionary Psychology and the Generation of Culture. Oxford U. Press (1992)

Darwiche, A., Pearl, J.: On the Logic of Iterated Belief Revision. Artificial Intelligence 89(1-2), 1–29 (1997), ftp://cobase.cs.ucla.edu/pub/stat_ser/R202.ps

Deacon, T.: Incomplete Nature: How Mind Emerged from Matter. Norton & Company (2011)

Doyle, J.: Inference and Acceptance. Computational Intelligence 10(1), 46–48 (1994), http://www.csc.ncsu.edu/faculty/doyle/publications/kyburg94.pdf

Friedman, N., Halpern, J.: Belief Revision: A Critique. Journal of Logic and Computation 18(5), 721–738 (2008)

Gärdenfors, P.: Knowledge in Flux: Modeling the Dynamics of Epistemic States. MIT Press (1988)

Gärdenfors, P. (ed.): Belief Revision, Cambridge Tracts in Theoretical Computer Science, vol. 29. Cambridge U. Press (1992)

Gärdenfors, P.: Conceptual Spaces: The Geometry of Thought. MIT Press (2000)

Gillies, A.: Two More Dogmas of Belief Revision: Justification and Justified Belief Change (2002), http://rci.rutgers.edu/~thony/just_and_revision.pdf

Hake, R.: Interactive-engagement Versus Traditional Methods: A six-thousand-student Survey of Mechanics Test Data for Introductory Physics Courses. American Journal of Physics 66(1), 64–74 (1998)

Hansson, B.: Infallibility and Incorrigibility. Philosophical Quarterly 18, 207–215 (2006)

Harper, W.: Rational Belief Change, Popper Functions and Counterfactuals. Synthese 30(1-2), 221–262 (1975)

Hestenes, D., Wells, M., Swackhamer, G.: Force Concept Inventory. Physics Teacher 30, 141–158 (1992)

Hilpinen, R.: Peirce's Logic. In: Handbook of the History of Logic, pp. 611–658. Elsevier BV (2004)

Hintikka, J.: Knowledge and Belief. Cornell U. Press (1962)

Hintikka, J.: Socratic Epistemology: Explorations of Knowledge-Seeking by Questioning. Cambridge U. Press (2007)

Hobbs, J., Stickel, M., Martin, P., Edwards, D.: Interpretation as Abduction. In: ACL 1988 Proceedings, 26th Annual Meeting of Association for Computational Linguistics, pp. 95–103 (1988)

von Humboldt, W.: Humbolt: On Language. Cambridge U. Press (1999)

Kapitan, T.: Peirce and the Structure of Abductive Inference. In: Houser, Roberts, Van Evra (eds.) Studies in the Philosophy of Charles Sanders Peirce, pp. 477–496. Indiana U. Press (1997)

Kasser, J.: Peirce's Supposed Psychologism. Transactions of the Charles S. Peirce Society XXXV, 501–526 (1999)

Kasser, J.: How Settled are Settled Beliefs in 'The Fixation of Belief'? Transactions of the Charles S. Peirce Society 47(2), 226–247 (2011)

Keeler, M.A., Pfeiffer, H.D.: Building a Pragmatic Methodology for KR Tool Research and Development. In: Schärfe, H., Hitzler, P., Øhrstrøm, P. (eds.) ICCS 2006. LNCS (LNAI), vol. 4068, pp. 314–330. Springer, Heidelberg (2006)

Keeler, M.A.: Revelator Game of Inquiry: A Peircean Challenge for Conceptual Structures in Application and Evolution. In: Priss, U., Polovina, S., Hill, R. (eds.) ICCS 2007. LNCS (LNAI), vol. 4604, pp. 443–459. Springer, Heidelberg (2007)

Keeler, M.A., Majumdar, A.: Revelator's Complex Adaptive Reasoning Methodology for Resource Infrastructure Evolution. In: Eklund, P., Haemmerlé, O. (eds.) ICCS 2008. LNCS (LNAI), vol. 5113, pp. 88–103. Springer, Heidelberg (2008)

Keeler, M.A.: Learning to Map the Virtual Evolution of Knowledge. In: Croitoru, M., Ferré, S., Lukose, D. (eds.) ICCS 2010. LNCS (LNAI), vol. 6208, pp. 108–124. Springer, Heidelberg (2010)

Kelly, K.: The Learning Power of Belief Revision. Recommended citation (1998), http://respository.cmu.edu/philosophy/389

Kyburg, H.: Probability and the Logic of Rational Belief. Wesleyan U. Press (1961)

Kyburg, H.: Believing on the Basis of Evidence. Computational Intelligence 10, 3–22 (1994)

Lehmann, D.: Belief Revision, Revised. In: Proceedings of the 14th International Joint Conference on Artificial Intelligence, vol. 2, pp. 1534–1540. Morgan Kaufmann (1995)

Levi, I.: Belief and Action. The Monist 48, 306–316 (1964)

Levi, I.: The Enterprise of Knowledge. The MIT Press, Cambridge (1980)

Levi, I.: Decisions and Revisions. Cambridge University Press (1984)

Levi, I.: Hard Choices: Decision Making Under Unresolved Conflict. Cambridge U. Press (1986)

Levi, I.: The Fixation of Belief and Its Undoing. Cambridge U. Press (1991)

Levi, I.: For the Sake of Argument: Ramsey Test Condtionals, Inductive Inference, and Nonmonotonic Reasoning. Cambridge University Press (1996)

Loviscach, J.: Vorlesungsaufzeichnungen auf YouTube. Herausforderungen, Werkzeuge, Erfahrungen (2012), http://www.youtube.com/watch?v=A34kAqyw8kM

Misak, C.: Peirce, Levi, and the Aims of Inquiry. Philosophy of Science 54(2), 256–265 (1987)

Mazur, E.: Peer Instruction: A User's Manual. Prentice Hall (1996)

Nayak, A., Pagucco, M., Peppas, P.: Dynamic Belief Revision Operators. Artificial Intelligence 146(2), 193–228 (2003)

Nesher, D.: Peirce's Essential Discovery: 'Our Senses as Reasoning Machines' Can Quasi-Prove Our Perceptual Judgments. Transactions of the Charles S. Peirce Society XXXVIII, 175–206 (2002)

Noschese, F.: What is pseudoteaching? (2001), http://fnoschese.wordpress.com/2011/02/21/pt-pseudoteaching-mit-physics/

Olsson, E., Enqvist, S. (eds.): Belief Revision Meets Philosophy of Science. Logic, Epistemology, and the Unity of Science, vol. 21. Springer (2011)

Park, W.: Belief Revision vs. Conceptual Change in Mathematics. In: Magnani, L., Carnielli, W., Pizzi, C. (eds.) Model-Based Reasoning in Science and Technology. SCI, vol. 314, pp. 121–134. Springer, Heidelberg (2010)

Pinker, S., Bloom, P.: Natural language and natural selection. Behavioral and Brain Sciences 13(4), 707 (1990)

Peirce, C.: The Fixation of Belief. Popular Science Monthly (1877)

Peirce, C.: Logical Machines. The American Journal of Psychology I, 165–170 (1887)

Peirce, C.: Collected Papers of Charles Sanders Peirce. In: Hartshorne, C., Weiss, P. (eds.) Elements of Logic, vol. II. Belknap Press of Harvard U. Press, Cambridge (1960)

Plotkin, H.: Evolution in Mind, An Introduction to Evolutionary Psychology. Harvard U. Press (1998)

Rott, H.: Two Dogmas of Belief Revision. The Journal of Philosophy 97(9), 503–522 (2000)

Santos, E.: On the Generation of Alternative Explanations with Implications for Belief Revsion. In: Proceedings of the Seventh Conference on Uncertainty in Artificial Intelligence, pp. 339–347. Morgan Kaufmann (1991)

Shermer, M.: The Believing Brain: How We Construct Beliefs and Reinforce Them as Truths. Henry Holt and Co., LLC (2011)

Subramanian, S., Mooney, R.: Combining Abduction and Theory Revision. Artificial Intelligence 92, 168 (1992)

Swami, V. (ed.): Evolutionary Psychology: A Critical Introduction. British Psychological Society & Blackwell Publishing Ltd. (2011)

Thagard, P.: The Best Explanation: Criteria for Theory Choice. The Journal of Philosophy 75(2), 76–92 (1978)

Thagard, P.: Computational Philosophy of Science. MIT Press (1988)

Thagard, P.: Concepts and Conceptual Change. Synthese 82(2), 255–274 (1990)

Thagard, P.: Conceptual Revolutions. Princeton U. Press (1992)

Van Fraassen, B.: The Pragmatics of Explanation. American Philosophical Quarterly 14(2), 143–150 (1977)

Walliser, B., Zwirn, D., Zwirn, H.: Abductive Logics in a Belief Revision Framework. Language and Information 14(1), 87–117 (2004)

Williams, M.-A., Rott, H.: Frontiers in Belief Revision. Kluwer Academic Publishers (2001)

Workman, L., Reader, W.: Evolutionary Psychology, an Introduction. Cambridge U. Press (2004)