Panel on belief change

Dagstuhl, August 11, 2005

This document gathers the panelists' contributions. The panelists were:

- Isaac Levi (Columbia University, New York, USA);
- Giacomo Bonanno (University of California at Davis, USA);
- Bernard Walliser (ENPC, Paris, France);
- Didier Dubois (IRIT-CNRS, Toulouse, France);
- Hans Rott (University of Regensburg, Germany);
- James Delgrande (Simon Fraser University, Vancouver, Canada);
- Jérôme Lang (IRIT-CNRS, Toulouse, France).

Contribution by Isaac Levi

The principle of categorial matching that has been invoked to mandate that transformations of theories by contraction removing some proposition h from a belief state K should include a specification of the entrenchment of the new belief state $K - h$ as well as the entrenchment of the initial state K masks a substantial assumption that ought to be subject to critical scrutiny and which personally would reject.

The assumption is that a belief state K and its entrenchment should be treated as a single unit for purposes of studying change in epistemic state.

To bring out the point, I took note of a similar issue that arises in the context of studying changes in state B of credal probability judgment relative to K due to expansion of K by adding datum E (consistent with K in conformity with the requirement that the change go by conditionalization via Bayes's theorem.

In this latter setting, the principle of categorial matching requires that the new state of credal probability judgment carry with it the specification of the rule of Bayesian updating. This rule would call for a specification of the conditional probability function $PE(x/y)$ relative to $K+E$ defined for all y consistent with $K + E$ where the conditional probability is understood to specify the new credal probability function were y added to $K + E$. This conditional probability is characterized differently than the conditional probability conditional on y

Dagstuhl Seminar Proceedings 05321
Belief Change in Rational Agents: Perspectives from Artificial Intelligence,
Philosophy, and Economics
http://drops.dagstuhl.de/opus/volltexte/2005/358

understood in terms of called off bets as in De Finetti, Ramsey, Shimony where $PE(x/y)$ is interpreted as the betting rate that would be fair for bets on x were these bets cancelled if y were false. The substantial assumption mandated by categorial matching in these cases is that the two interpretations of conditional probability are equivalent.

This becomes apparent when subjective or credal probability is taken to be controlled by a rule representable by a function from potential states of full belief in K to states of credal probability belonging in B. Call such a function $F : K \to B$. Traditional Bayesians from Bayes to H.Jeffreys and R. Carnap all presupposed that such a function satisfied a constraint that stipulated that $F(K + E)(x/y) = F(K)(x/yE)$. I have called a generalization of such a function to cases where credal probability is indeterminate confirmational conditionalization.

In cases where a change in state of full belief or "evidence" K to $K + E$ takes place without any change in F, we have what I have called temporal credal conditionalization. This corresponds to the usual updating according to conditionalization via Bayes theorem. It meets the requirement of categorical matching. As long as the confirmational commitment does not change, the two senses of conditional credal probability coincide. But satisfying categorical matching is not mandated in all cases. Confirmational commitments are revisable independent of changes in states of full belief. The usual view of conditionalization fails to countenance this possibility. Recognizing the revisability of confirmational commitments entails abandoning categorial matching as a condition on changing credal states.

I suggest that a similar situation arises in the case of contraction. As long as certain "demands for information" are held fixed, contraction of K by removing h uniquely determines the contraction and the new entrenchment. (See *Mild Contraction* for details.) But the demands for information can change independently of whether the contraction removing h takes place. A new entrenchment will emerge but it will not be a function of the initial state of full belief, entrenchment and input specification of what is to be removed as devotees of the principle of categorial matching seem to think.

There are plenty of other examples where factors that may be independently changeable are packaged together and change in the pair is taken as the focus of study without any serious consideration of their separability.

My doubts about invoking categorial match concern this question-begging practice.

Contribution by Giacomo Bonanno

1. Remarks on the notion of information

The theories of belief revision that have been discussed in this workshop deal with the transition from a belief state to a new belief state in response to a piece of information received by the agent. Information is typically treated as veridical and the success axiom is assumed, requiring information to be believed. None of the talks have addressed the issue of what can/must be treated as information and under what circumstances information can be safely taken to be correct. Even reputable sources of information can sometimes convey false information. A clear illustration of this can be found in the following newspaper article (The Sacramento Bee, September 1, 2000): "Mark J. made a big bet in mid-August that Emulex shares would decline, federal prosecutors say. Instead they soared, leaving him with a paper loss of almost $100,000 in just a week. So J. took matters into his own hands. [...] On the evening of August 24, he sent a fake press release by e-mail to Internet Wire, a Los Angeles service where he had previously worked, warning that Emulex's chief executive had resigned and its earnings were overstated. The next morning, just as financial markets opened, Internet Wire distributed the damaging release to news organizations and Web sites. An hour later, shareholders in Emulex were $2.5 billion poorer. And J. would soon be $240,000 richer. [...] The hoax [...] was revealed within an hour of the first news report and Emulex stock recovered the same day. Still, investors who [believed the fake news release and] panicked and sold their shares, or had sell orders automatically executed at present prices, are unlikely to recover their losses".

A related point concerns iterated belief revision. Insistence on the success axiom seems hard to justify in an iterated context. If the same source, over time, releases contradictory pieces of information, why should the agent continue to treat the information at face value? Shouldn't the credibility of the source be doubted and perhaps the information itself be completely dismissed? In interactive contexts one needs to study the incentives that agents have to convey information. There may be situations where it is in the best interest of an agent to convey only partial or even false information. Realizing this, the other agents cannot rationally treat information as reliable. Aside from strategic considerations, when the source of information is another agent, all the informed party learns is that the party that claims A to be the case merely believes A to be the case. Thus there are situations where information perhaps is best modeled as the (justified) belief of another agent, so that belief revision perhaps can be modeled within the framework of interactive beliefs.

2. Remarks on the AGM framework

The AGM axioms are more about belief revision policies than about actual belief revision. The AGM theory insists on specifying how the agent would revise his beliefs in every possible circumstance, that is, if faced with every conceivable piece of information (even inconsistent information). This is reminiscent of the notion of strategy in a dynamic game. A player's strategy specifies a plan of action for every conceivable situation in which the player might have to move (even those situations that the player knows will not arise, e.g. because his own planned moves will rule them out)). The role of a strategy in a game is not so much to model the planning process of the player himself but rather a way of expressing the beliefs in the minds of the other players about how the player under consideration would act in different situations. In the single-agent belief revision framework, on the other hand, it is not clear why one should insist on a complete revision plan rather than investigate principles guiding actual belief revision. From a practical point of view, the most we can observe is how an agent responds to some sequence of pieces of information. From that it may be not be possible, or even interesting, to determine how the agent would have reacted to a different sequence of pieces of information. Does the notion of rationality really require a complete belief revision policy?

Contribution by Bernard Walliser

Belief revision and experimentation

In epistemic logics or Artificial Intelligence, belief revision gives rise to a huge theoretical work, but with little empirical testing (even if a few experiments were done after the preliminary work of Kahneman and Tversky on the Bayes rule). In this respect, the situation is similar to decision and game theory ten years ago, where empirical validation was not very much practized and even badly considered.

However, such an empirical investigation looks rather easy to implement in laboratory, especially with the help of cognitive psychologists. Some experimental problems, already suggested or specially imagined, can be proposed to different populations. A well-known difficulty is that people do not react in the same way in abstract situations than in more concrete ones. The problem of extrapolating from laboratory to real life is also to consider.

Empirical investigation depends on what can be observed. It is possible to test directly specific belief revision rules by observing simple enough beliefs (like probabilities). Some belief revision axioms may be directly tested by observing beliefs too. Conversely, revision rules may be tested only indirectly by observing decisions taken by an agent. Finally, like in neuroeconomics, tests can be made by observing the activity of the brain.

Usual experimentation is projective, since the scientist analyzes what are the consequences of a given principle and checks if these testable consequences are confirmed or refuted. Less often, experimentation is inductive, since the scientist tries to infer revision rules from observed regularities. Concretely, both ways are used simultaneously, projection concerning a class of models and induction making precise some parameters.

The problem is to evaluate if some biases from existing theories are clearly systematic and cannot be attributed to experimental conditions. It is only in that case that it is possible to persuade the modeller that he has to refute a revision rule, hence some underlying axioms, and to replace them by others.

However, it is always possible to consider that the rules are only contextual, i.e. are valid for some specific problems and environments.

Contribution by Didier Dubois

I see two problematic issues or misunderstandings about belief revision and merging:

1. The fact that the main approaches are essentially syntax-independent but that most people continue to discuss the postulates inside a syntactic language, adding a syntax-independence axiom. This method looks artificial to me, and can be misleading. It makes things esoteric, often.

2. The fact that people never really discuss what an input is with respect to an agent's epistemic state. And what actually an epistemic state contains. There are certainly more than one anwswer to it, but this answer totally conditions what iterated revision is and if it makes sense or not.

The first issue was not too much discussed. I elaborate on the second one. One view (see [DFP04]) can be that an epistemic state is made of

- a plausibility ordering of states of nature describing what is generally normal and what is less normal in the world

- a set of facts considered true for the particular world of interest w (hence consistent with one another). This particular world is NOT supposed to evolve (can be a past case to elucidate).

- a set of beliefs about the particular world of interest. These beliefs are induced by the two other items.

Under this view, the Gärdenfors notation $K * A$ for the result of revising K by A means the following: An external observer sees the agent's beliefs about w evolve from K to $K * A$ when the agent gets the information A about the particular world of interest. How the agent buids up $K * A$ is by focusing the ordering of states of nature on the sets of A-worlds. So plausible beliefs are inferred from the observations and the generic knowledge. Then it has consequences:

1. You never need K to derive $K * A$, you only need $*$ (= the ordering) and A. So the notation $K * A$ is just misleading;

2. The belief revision step leaves the ordering of states unchanged. This is because inputs and the plausible ordering deal with different matters, resp the particular world of interest, and the class of world the plausible ordering refers to.

3. While beliefs states seem to evolve as if iterated belief revision would take place, $K * A * B$ is really obtained by gathering the available observations A

and B and inferring plausible beliefs from then. Again we do not compute $K * A * B$ from $K * A$. But $K * A * B = K * (A \wedge B)$ (not obtained from K), with the proviso that A and B should be consistent. So a merging of observation is operated.

4. If observations A, B are inconsistent then the above view collapses, because it means that some of the facts are wrong. Then the merging step must be non-trivial, and carried our in order to come up with a consistent context from which to infer plausible beliefs.

5. Under this view the belief revision problem (absorbing in observations) becomes totally different from the problem of revising the plausible ordering of states of nature. In particular it makes no sense to revise an ordering by a formula, for instance.

A radically different view is to consider that an epistemic state is made of uncertain evidence about a particular world of interest (static, again). Then the plausibility ordering is no longer like a statistical distribution, it is like a "belief function" in the sense of Shafer's mathematical theory of evidence. It gathers the past uncertain observations obtained so far, and the new observations (which can be unreliable, uncertain) have the same status as the plausibility ordering. So here it makes sense to speak of iterated revision, if the plausibility orsering and the new observation are merged in a non-commutative way. But this is a matter of option. If we postulate that all uncertain observations play the same role, this is again a symmetric and possibly associative merging process.

It is not clear what Gärdenfors theory is talking about : the first view or the second one. I tend to believe it is the first one, except that, in the AGM theory, the plausibility ordering is a by-product of the postulates (it says: if from the outside, an agent's beliefs seem to evolve according to the postulates, then it is as if there were a plausibility ordering that drives the belief flux). I find it more natural to use the plausibility ordering as a primitive explicit ingredient and take an insider point of view, rather than observing beliefs change from the outside.

[DFP04] D. Dubois, H. Fargier and H. Prade. Ordinal and probabilistic representations of acceptance. *J. Artificial Intelligence Research*, 22, 23-56, 2004.

Contribution by Jim Delgrande

To my mind the two most pressing issues in belief change are:

1. formalizing iterated belief change, and

2. determining a taxonomy of belief change operators.

1. Iterated belief change

The major issues here are: how to appropriately model iterated change and determining the appropriate properties (and for which circumstances) of iterated change. (Since this divides roughly into semantics and syntax, this distinction doesn't mean a whole lot!)

For revision and contraction, the favourite model remains ordinal conditional functions. But at present, there are no agreed upon approaches using OCF's (although Darwiche/Pearl is held as a standard for comparison), and succeeding proposals each seem to be prone to counterexamples, typically illustrating some sort or other of a 'drowning effect'. More recently there have been general distance-based approaches that have been developed. A nice feature of these approaches is that they separate revising beliefs from revising the background epistemic state (given by a set of worlds/interpretations and a distance function on these worlds/interpretatins). Given a lack of agreement concerning basic properties of iteration, clearly there is more work to be done here.

2. Determining a taxonomy of belief change operators

It is, as yet, not fully clear just what the appropriate operators are that constitute belief change, nor what the appropriate properties of these operators are, nor what problems each is best suited for. Since research in belief change is not as advanced as other areas (nonmonotonic reasoning comes to mind as an example) it is not surprising that an overall 'lay of the land' is only now starting to emerge.

For example, merging is currently a hot topic in belief change, and there is nice work going on in proposing and examining different types of merging operators. However (IMO!) the landscape isn't yet clear with regard to merging operators; it isn't clear for example whether there are other major types of merging operators that have yet to be developed. For this last point, it may be worthwhile looking at work in topology, say, or computational geometry, which deal with notions of distance, closeness, centres of mass, etc., and via (say) work in computational geometry try to come up with an overarching theory of merging based on these given notions of distance, etc.

As well, even for 'established' operators, it isn't clear that accepted notions are in fact sound. Hans Rott's "Two Dogmas" paper is a very nice example of work that questions underlying assumptions in belief change. Here's another example: in belief revision, it is commonly accepted that revision concerns changes in an agent's beliefs about a static world. As well, it is accepted that newer information should take precedence over less recent information. But this is problematic. Consider a past party that you missed (since you were on vacation) but on coming home you find a sequence of messages on your answering machine.

Person A says: "I have it on good authority that John was there".

Person B says: "I have it on good authority that if John was there then so way Mary".

Person C says: ”I have it on good authority that John was not there”.

Person B phoned the day after person A, and person C the day after person B. Assume that these people don’t know each other.

The initial knowledge base can be represented by TRUE (i.e. nothing is known about the party). Standard accounts based on AGM revision indicate that the revisions should be carried out as $((K * J) * J \rightarrow M) * \neg J$. But clearly this doesn’t make sense: we have 3 reports from the party and no reason to give one report more credence than another. What should be done in this case then remains open: perhaps K should be revised separately by each piece of information and the results merged, or perhaps the items should be merged and then K revised by the result, or perhaps something entirely different should be done. In any case, we’re now back to issue 1, iterated belief change!

Contribution by Hans Rott

In the beginning (the 1980s) there was the classical AGM paradigm, with its obvious limitations. The 1990s were the decade in which people realized that the problem of iteration - that not had not been treated by AGM - is highly non-trivial and has many solutions, none of which is suitable for all purposes. It turned out that the kinds of considerations motivating the one-step revision case, like ”informational economy” (which is much overrated anyway), are not sufficient to deal with problem iterations. Two general conclusions emerged.

First, it is in any case inadequate to identify a belief state with the set of beliefs held by the agent. Conditionals may hold the key for the dynamics of belief, but the ramifications of the impossibility theorems surrounding the Ramsey test are still not fully understood. Belief states need to encompass some sort of revision-guiding structure, most frequently captured in the form of a preference relation or a choice function, but the syntactical structure of a belief base is often equally important. Belief states have two functions at least: They should allow us to retrieve the current belief set, and revision functions should operate on belief states directly.

Second, we need to sort out carefully the motivating principles behind the many particular approaches to iterated belief change, study their interaction and assess their performance in various application contexts. I think that there is a need for a methodology of belief change: Which methods are good for which purposes, and for what reasons? It is important to have a deepened understanding of what can be achieved with purely qualitative or relational approaches, and in which contexts there is no way of getting around employing numbers, in order to represent weights of or distances between beliefs. Quantitative approaches like probability theory or ranking functions are very powerful, but we must be clear about the fact that meaningful numbers are quite hard to come by.

Studies of belief transformations should be better linked with studies of belief formation. By, this I mainly mean the drawing of inferences from, or more generally, the processing of a data base that ultimately yields a well-balanced belief state. Various possible ”logics” that can be used will give quite different perspectives on the topic. For instance, why change your beliefs in the first place, if you have a paraconsistent logic that can take care of the inconsistencies to be found in your database? Why bother about sophisticated change operations if you have a sophisticated nonmonotonic logic that helps you extend scarce information bases?

An fascinating field that has come to the fore in the last years is that of judgment aggregation (also discussed under the headings ”discursive dilemma”, ”doctrinal paradox” and ”decisions on multiple propositions” – see

http://personal.lse.ac.uk/LIST/doctrinalparadox.htm

– which should have a much more intimate relationship with belief revision theories than it currently has.

There has been a proliferation of accounts of belief revision over the last 20 years. To be sure, this diversity is good. But from my point of view, the philosophical foundations of belief revision theories are not as well understood as their technical properties. I am optimistic, however, that there is much progress in this field ahead of us. Of the major trends and developments of the last years, I find the field of belief fusion or belief merging particularly important, both from a conceptual and a mathematical perspective. Impressive work has already been done, and there is still much to be explored. Further connections with models for practical rationality like social choice theory or game theory will lead to exciting new developments.

Contribution by Jérôme Lang

Belief change can be thought of as the construction of an adequate system for reasoning about a possibly uncertain and possibly evolving world. Quoting (Friedman and Halpern, 96), before discussing about postulates and representation theorems we should first make it clear what kind of knowledge we are representing and processing. This methodology is also at work in Sandewall's book (Sandewall, 94) where most logical approaches for reasoning about change existing at that time (1994) are evaluated with respect to the underlying assumptions about the world ("ontological' specialities) and the agent's beliefs ('epistemological' specialities). Since then, several new belief change paradigms have arised, many new papers have been written, so there is a need for a reactualization of Sandewall's taxonomy. Yet, the important thing is that in most papers, no effort is devoted to make it clear what the ontological and epistemological assumptions underlying the approach are.

In this short note I focus on *purely inertial* (more often called *static* – I prefer "purely inertial" that I find less ambiguous) approaches to belief change. Pure inertia is characterized by the fact that *the state of the world remains constant* as time goes on. (When I say "the world" I of course mean the abstraction of the world, restricted to the propositions of interest.) This is equivalent to saying that

a. the agent cannot perform any action intended to make the world evolve;

b. no exogeneous actions can occur. (Exogeneous actions are either events – whose occurrence is nature-driven – or actions performed by other agents.)

Under this strong assumption, the pieces of informations available to the agent describe some properties (possibly incomplete and/or partially unreliable) of the actual state of the world s^*. Without much loss of generality we can call these pieces of information "observations".

Stated differently, assume that we have a finite time scale $1, \ldots, n$ and a sequence of observations $\langle \alpha_1, \ldots, \alpha_n \rangle$, where α_i has been obtained at time i; then a trajectory (= sequence of states) $\tau = \langle s_1, \ldots, s_n \rangle$ is possible only if $s_1 = s_2 = \ldots = s_n (= s^*)$. In addition to these observations α_i we may have a data structure K expressing *prior belief* about the state of the world.

In brief, a belief change operator for strongly inert domains maps a prior belief structure K and a sequence of observations $\langle \alpha_1, \ldots, \alpha_n \rangle$, all (including K) referring to the *same* state s^* of the world to a belief structure (again, this formulation is left vague on purpose), denoted by

$$*(K, \alpha_1, \ldots, \alpha_{n-1})$$

and expressing the agent's beliefs after taking account all these pieces of information. In other terms, one has to *merge* different pieces of information about the same state of affairs so as to get an *aggregated belief.* Such a process is usually called *belief merging.*

A first special case of belief merging arises when all pieces of information α_i have the same reliability (and K is empty). This gives raise to *commutative merging*, as studied by several authors. A second special case of belief merging arises when the pieces of information are ordered in such a way that any single information of higher priority takes precedence over any subset of pieces of information of lower priority. This will be referred to as *prioritized merging.*

Now, iterated belief revision, as studied in many papers, addresses the issue of aggregating several pieces of information about the same state of affairs – and is therefore a specific case of belief merging. it is however based on an extra assumption – namely, that the pieces of information (observations) come in such an order that any observation has precedence over any subset of observations that came earlier. Therefore, I argue that *iterated revision is prioritized merging*, where priority depends directly from the time when the observations were performed, following the *priority to recency* principle (the sooner, the better).

What does "priority to recency" really mean? A wrong idea that is sometimes informally exposed is that the older a piece of information, the weakest, because it is more likely that it failed to persist than a more recent one. This would be perfectly ok in a dynamic framework, where events can occur (even rarely) and induce unpredicted and possibly unobserved changes in the world. However, this is meaningless in a purely inertial framework where the state of the world does not change. In this case, the order in which the observations are made is not really significant in itself. It is not difficult to imagine that we might have got the very same observations in a different order, the reliability of each observation remaining the same. In this case, the order in which pieces of information are taken into account does not generally reflect time, but only priority: I first collect my n pieces of information, I evaluate their priority (reflecting their reliability) and then I may merge them using an iterated revision process, taking them in the order induced by reliability (which has nothign to do with the time when they were collected). This is not what most papers about IBR intend to do (just look at the examples): there, pieces of informations are taken into account as soon as they arrive, and priority to recency is applied. I cannot see any reason why priority to recency would be plausible (why should always the last observation be considered more reliable than previous ones, since they all relate to the same state of affairs?) Right, strange situations where it makes sense can be imagined: someone moving forward to on object and gets closer and closer to it, or different sensors with inceasing accuracy being used of, but these remain very specific situations which would probably not deserve this huge literature!

What is the status of the acceptance postulate? Since we want a non-trivial handling of contradictions, observations have to be considered non-reliable by default – the main issue then consists in identifying the wrong observations. No problem for considering some observations as fully reliable (and should be never given up) as in merging with constraints (e.g. Konieczny and Pino-Perez, 99) as well as in standard (non-iterated revision, which is here seen as merging a defeasible piece of information and an undefeasible one ("revision is prioritized expansion", to quote A. Bochman). However, in the case of iterated revision / prioritized merging, requiring acceptance, that is, considering the last observation as fully reliable, contradicts the possibility that a later and more entrenched observation might leads to give it up.

However, in all cases there is a principle of *default acceptance* that leads to accept observations as soon as they are jointly consistent. Didier calls this "closure by optimism", because this principle implicitly relies on the assumption that failure are rare (normally, observations are truthful).

And now, what about all these nice postulates? If we accept this view of iterated revision as prioritized merging, do the postulates for iterated revision still make sense? Technically speaking, they remain unchanged, but they have to be reinterpreted (replacing time by reliability/priority). An important difference is that with the temporal interpretation, whereas it makes sense to discuss the relationships between $\star(\alpha_1, \ldots, \alpha_n)$ and $\star(\alpha_1, \ldots, \alpha_n, \beta)$ – in fact, this is what all postulates are about – it makes much less sense to discuss the relationship between $\star(\alpha_1, \ldots, \alpha_n)$ and $\star(\alpha_1, \ldots, \alpha_k, \beta, \alpha_{k+1}, \ldots, \alpha_n)$ for $k < n$. In the priority interpretation both make sense: there is no reason why we should consider pieces of information in the increasing priority order, and it is relevant to discuss the impact of adding a new piece of information *of any priority* to a collection of pieces of information. Further work would then consist in considering the postulates (and the representation theorems) one after the other, and to understand what they really mean – having in mind that iterated revision is interpreted as prioritized merging.