A Counterexample to Six Fundamental Principles of Belief Formation

Hans Rott
Department of Philosophy
University of Regensburg
93040 Regensburg
Germany

hans.rott@psk.uni-regensburg.de

Abstract

In recent years there has been a growing consensus that ordinary reasoning does not conform to the laws of classical logic, but is rather nonmonotonic in the sense that conclusions previously drawn may well be removed upon acquiring further information. Even so, rational belief formation has up to now been modelled as conforming to some important principles that are classically valid. The counterexample described in this paper shows that a number of the most cherished of these principles should not be regarded as valid for commonsense reasoning. An explanation of this puzzling failure is given that transfers insights from the theory of rational choice to the realm of belief formation.

1 Introduction

Part of the cognitive state of a person is characterized by the set of her beliefs and expectations. By the term 'belief formation', we will refer to two different kinds of processes. The first one is that of drawing *inferences* from a given set of sentences. Ideally, as a result of this process the reasoner arrives at a well-balanced set of beliefs in reflective equilibrium, sometimes referred to as the 'belief set'. The second process is that of readjusting one's belief set in response to some perturbation from outside ('belief transformation' might be a better term in this case). We will consider two subspecies of belief change that may be triggered by external perturbations. If there is some cognitive 'input', a sentence to be accepted, we speak of a belief *revision*. If the reasoner has to withdraw one of her beliefs, without accepting another belief in its place, she performs a belief

contraction. We will be dealing with the three topics of inference, revision and contraction in turn. For each of them, we consider two qualitative principles that have up to now been regarded as very plausible ones. We shall then tell a story with a few alternative developments which is intended to show that all of these principles fail. They do not fail due to the contingencies of some particular system that is being proposed, but indeed as norms for good reasoning. Based on recent reconstructions of belief formation in terms of the theory of rational choice, we give an explanation of why these principles fail. It turns out that well-known problems of this very general theory transfer to the special case where the theory is applied in operations of belief formation. In fact, it will turn out that this special case has features that block one of the standard excuses for the problem at hand. We end up in a quandary that poses a serious challenge to any future conception of belief formation procedures.

For a long time, the notion of inference has been thought to be identical with the notion of logical consequence or deduction. Partly as a result of the problems encountered in research in artificial intelligence and knowledge representation during the 1960s and 1970s, however, logicians have come to realize that most of our reasoning proceeds on the basis of incomplete knowledge and insufficient evidence. Implicit assumptions about the normal state and development of the world, also know as expectations, presumptions, prejudices or defaults, step in to fill the gaps in the reasoner's body of knowledge. These default assumptions form the context for ordinary reasoning processes. They help us to generate conclusions that are necessary for reaching decisions about how to act, but they are retractible if further evidence arises. Thus our inferences will in many contexts be *defeasible* or *non-monotonic* in the sense that an extension of the set of premises does not generally result in an increase of the set of legitimate conclusions. This, however, does not mean that the classical concept of logical consequence gets useless, or that our reasoning gets completely irregular. For the purposes of this paper, we can in fact assume[1] that the set of our beliefs (and similarly: the set of our expectations) is consistent and closed with respect to some broadly classical consequence operation Cn. This combined notion of logical coherence (consistency-cum-closure) may be viewed as a constraint that makes the processes of inference and belief change a non-trivial task.[2]

2 Six fundamental principles of belief formation

[1] Along with Stalnaker (1984, p. 82) and Dennett (1987, p. 21), as well as the majority of the more technical literature mentioned below.

[2] What has been said in this little paragraph takes the position that the nonmonotonicity of commonsense reasoning is an effect of a certain way of using classical logic, rather than a result of applying some irreducibly non-classical, ampliative inference operation. For the latest state of this debate, see Morgan (2000) and Kyburg (2001). The counterexample below is independent of any particular philosophical stand in this matter.

2.1. In the last two decades, a great number of systems for non-monotonic reasoning have been devised that are supposed to cope with the newly discovered challenge.[3] Many classical inference patterns are violated by such systems, but it is equally important to keep in mind that quite a number of classical inference patterns do remain valid. Let us now have a look at two properties that have usually been taken to be constitutive of sound reasoning with logical connectives like 'and' and 'or' even in the absence of monotonicity.

First, if the premise x allows the reasoner to conclude that y is true, then y may be conjoined to the premise x, without spoiling any of the conclusions that x alone permits to be drawn. This severely restricted form of the classical monotony condition is usually called *Cumulative Monotony* or *Cautious Monotony*:

(1) If y is in $Inf(x)$, then $Inf(x) \subseteq Inf(x \wedge y)$

Here $Inf(x)$ denotes the set of all conclusions that can be drawn if the premise is x. The reasoner may in fact possess an arbitrary finite number of premises which are conjunctively tied together in x. More importantly, $Inf(x)$ is meant to denote what can be obtained if x is *all* the information available to the agent.

Secondly, if a reasoner wants to know what to infer from a disjunction $x \vee y$, she may reason by cases. She will consider first what would hold on the assumption of x, and then consider what would hold on the assumption of y. Any sentence that may be inferred in both of these cases should be identifiable as a conclusion of an inference starting from $x \vee y$. This is the content of a condition called *Disjunction in the Premises*:

(2) $Inf(x) \cap Inf(y) \subseteq Inf(x \vee y)$

Cumulative Monotony and Disjunction in the Premises hold in most defeasible reasoning systems that have been proposed since non-monotonic logic came into being, and especially in those systems that are semantically well-motivated. There is one important and striking exception. Reiter's (1980) seminal system of Default Logic violates both Cumulative Monotony and Disjunction in the Premises. However, no advocate of Reiter's logic has ever argued that Cumulative Monotony and Disjunction in the Premises *should* be violated. These violations have usually been taken to be defects of the system that need to be remedied.[4] Conditions (1) and (2) have never lost their normative force.

2.2. Let us now turn to the revision of belief sets in response to new information. If someone has to incorporate a conjunction $x \wedge y$, she has to accept both x and y. One idea how to go about revising by the conjunction is to revise first with x. If it so happens that y is accepted in the resulting belief set, then one should be sure that every belief contained in this set is also believed after a revision of the original belief set by $x \wedge y$. In the following, $B * x$ denotes the set of beliefs held

[3] For a an excellent survey of the logical patterns underlying nonmonotonic reasoning, see Makinson (1994).

[4] See for instance the discussions in Brewka (1991), Giordano and Martelli (1994) and Roos (1998).

after revising the initial belief set B by x (and likewise for the input sentence $x \wedge y$).

(3) If y is in $B * x$, then $B * x \subseteq B * (x \wedge y)$

Another approach to circumscribing the result of a revision by the conjunction $x \wedge y$ is to revise first with x, and then to just add y set-theoretically and take the logical consequences of everything taken together. This is not always a good idea, since y may be inconsistent with $B * x$, and thus the second step would leave us with the inconsistent set of all sentences. But even if we may end up with too many sentences, this strategy seems unobjectionable if it is taken as yielding an upper bound for the revision by a conjunction. This is the content of principle

(4) $B * (x \wedge y) \subseteq Cn((B * x) \cup \{y\})$

2.3. Finally, we consider the removal of beliefs. Here again, we focus on upper and lower bounds of changes with respect to conjunctions. If a person wants to remove effectively a conjunction $x \wedge y$, she has to remove at least one of the conjuncts, that is, either x or y. So if the second conjunct y is still retained in the result of removing the conjunction, what has happened is exactly that the first conjunct x has been removed. We will be content here with a weaker condition that replaces the identity by an inclusion. Here and elsewhere, $B \dot{-} x$ denotes the set of beliefs that are retained after withdrawing x from the initial belief set B (and likewise for the case where $x \wedge y$ is to be discarded):

(5) If y is in $B \dot{-} (x \wedge y)$, then $B \dot{-} (x \wedge y) \subseteq B \dot{-} x$

Another approach to circumscribing the result of a contraction with respect to the conjunction $x \wedge y$ is to consider first what would be the result of removing x, and then consider what would be the result of removing y. It is not always necessary to take into account both possibilities, but doing so should certainly be suitable for setting a lower bound. Any sentence that survives both of these thought experiments should surely be included in the result of the removal of $x \wedge y$. This is the content of principle

(6) $B \dot{-} x \cap B \dot{-} y \subseteq B \dot{-} (x \wedge y)$

Principles (3) – (6) have been endorsed almost universally in the literature on belief revision and contraction. The classic standard was set by Alchourrón, Gärdenfors and Makinson (1985). Conditions (4) and (6) are the seventh of their eight 'rationality postulates' for revision and contraction, conditions (3) and (5) are considerably weaker—and thus considerably less objectionable—variants of their eighth postulates.[5] There exist sophisticated 'translations' between operations of nonmonotonic inference, belief revision and removal which show that notwithstanding different appearances, conditions (1), (3) and (5) are essentially different sides of the same (three-faced) coin, as are conditions (2), (4) and (6).[6]

[5] For comprehensive treatments of this topic, see Gärdenfors (1988), Gärdenfors and Rott (1995) and Hansson (1999).

[6] See Makinson and Gärdenfors (1991) and Rott (2001, chapter 4).

Some methods of belief formation suggested in the literature violate one or the other of the six principles. Nevertheless, it is fair to say that these principles have retained their great intuitive appeal and have stood fast up to the present day as norms to which all good reasoning is supposed to conform.

But this is wrong, or so I shall argue. The following section presents an example that shows, I think, that *not a single one* of the six principles listed above ought to be endorsed as a valid principle of rational belief formation.

3 The counterexample

The story goes as follows. A well-known philosophy department has announced an open position in metaphysics. Among the applicants for the job there are a few persons that Paul, an interested bystander, happens to know. First, there is Amanda Andrews, an outstanding specialist in metaphysics. Second, we have Bernice Becker, who is also definitely a very good, though not quite as excellent a metaphysician as Andrews. Becker has in addition done some substantial work in logic. A third applicant is Carlos Cortez. He has a comparatively slim record in metaphysics, but he is widely recognized as one of the most brilliant logicians of his generation.

Suppose that Paul's initial set of beliefs and expectations includes that neither Andrews nor Becker nor Cortez will get the job (say, because Paul and everybody else thinks that David Donaldson, a star metaphysician, is the obvious candidate who is going to get the position anyway). Paul is aware of the fact that only one of the contenders can get a job.

3.1. Consider now three hypothetical scenarios, each of which describes a potential development of the selection procedure. The scenarios are not meant as describing consecutive stages of a single procedure. At most one of the potential scenarios can turn out to become real. In each of these alternative scenarios, Paul is genuinely taken by surprise, because he learns that one of the candidates he had believed to be turned down will—or at least may—be offered the position. (Donaldson, by the way, has told the department that he has accepted an offer from Berkeley.) To make things shorter, we introduce some abbreviations. Let the letters a, b and c stand for the sentences that Andrews, Becker and Cortez, respectively, will be offered the position.

Scenario 1. The dean tells Paul in confidence that it has already been decided that either Andrews or Becker will be appointed. This message comes down to supplying Paul with the premise $a \vee b$. Given this piece of information, Paul concludes that Andrews, being the better metaphysician, will get the job. He also infers that all the other candidates are going to return empty-handed.

Scenario 2. In this scenario the dean tells Paul that it has been decided that either Andrews or Becker or Cortez will get the job, thus supplying him with the premise $a \vee b \vee c$. This piece of information triggers off a rather subtle line of

reasoning. Knowing that Cortez is a splendid logician, but that he can hardly be called a metaphysician, Paul comes to realize that competence in logic is regarded as a considerable asset by the selection committee. Still, Paul keeps on believing that Cortez will not make it, because his credentials in metaphysics are just too weak. Since, however, logic appears to contribute positively to a candidate's research profile, Paul concludes that Becker, and not Andrews, will get the job.

This qualitative description should do for our purposes, but for those who prefer the precision of numbers, the following elaboration of our story can be given. Suppose that the selection committee has decided to assign numerical values in order to evaluate the candidates' work. Andrews scores 97 out of 100 in metaphysics, but she has done no logic whatsoever, so she scores 0 here. Becker scores 92 in metaphysics and a respectable 50 in logic. Cortez scores only 40 in metaphysics, but boasts of 99 in logic. In scenario 1, Paul takes it that metaphysics is the only criterion, so clearly Andrews must be the winner in his eyes. But in scenario 2, Paul gathers that, rather unexpectedly, logic has some importance. As can easily be verified, any weight he may wish to attach to the logic score between $^1/_{10}$ and $^1/_2$ (with metaphysics taking the rest) will see Becker ending up in front of both Andrews and Cortez.

Scenario 3. This is a very surprising scenario in which Paul is told that Cortez is actually the only serious candidate left in the competition. There is little need to invest a lot of thinking. Paul accepts c in this case.

Let us summarize the scenarios as regards the conclusions Paul would draw from the various premises that he may get from the dean of the faculty. In scenario 1, Paul infers from $a \vee b$ that a and $\neg b$ (along with $\neg c$ and $\neg d$ which we will not mention any more). In scenario 2, he infers from $a \vee b \vee c$ that $\neg a$ and b. In scenario 3, he infers from c that $\neg a$ and $\neg b$.

Now we first find that this situation does not conform to Cumulative Monotony. Substitute $a \vee b \vee c$ for x and $a \vee b$ for y in (1). Even though Paul concludes that $a \vee b$ is true on the basis of the premise $a \vee b \vee c$, it is not the case that everything inferable from the latter is also inferable from $(a \vee b \vee c) \wedge (a \vee b)$ which is equivalent with $a \vee b$. Sentences $\neg a$ and b are counterexamples.

Secondly, the example at the same time shows that Disjunction in the Premises does not hold. Take (2) and substitute $a \vee b$ for x and c for y. Then notice that $\neg b$ can be inferred both from $a \vee b$ and from c, but it cannot be inferred from $a \vee b \vee c$.

Summing up, even though Paul's reasoning is perfectly rational and sound, it violates both Cumulative Monotony and Disjunction in the Premises.

3.2. Let us then turn to the dynamics of belief. The case of potential revisions of belief is very similar to the case of default reasoning. What we have so far considered as the set of all sentences that may be inferred from a given premise x, will now be reinterpreted as the result of revising a belief set by a new piece of information.

This is best explained by looking at the concrete case of the selection procedure.

Paul's initial belief set B contains $\neg a$, $\neg b$, $\neg c$ and d (among other things). Paying attention to the fact that the structure of (3) is very similar to the structure of (1), we can re-use the above argument. If Paul's set of initial beliefs and expectations is revised by $a \vee b \vee c$, then the resulting belief set includes $a \vee b$ (because it includes a). However, the revised belief set $B * (a \vee b \vee c)$ is not a subset of the belief set $B * ((a \vee b \vee c) \wedge (a \vee b)) = B * (a \vee b)$, as is borne out by sentences like $\neg a$ and b. Thus (3) is violated.

In principle (4), substitute $a \vee b \vee c$ for x and $a \vee b$ for y. Then the left-hand side is changed to $B * (a \vee b)$, while the right-hand side consists of the set of all logical consequences of $B * (a \vee b \vee c)$ and $a \vee b$ taken together. Since $a \vee b$ is already included in $B * (a \vee b \vee c)$, we need only consider this latter set. But as we have by now seen several times, the two revised belief sets just mentioned cannot be compared in terms of the subset relation. So (4) is violated.

3.3. For the consideration of belief contractions, we have to change our story slightly. Suppose now that in the different scenarios Paul may be going through, the dean does not go so far as to tell him that Andrews or Becker (or Cortez) *will* get the offer, but only that one of them *might* get the offer. Paul's proper response to this is to *withdraw* his prior belief that none of Andrews and Becker (and Cortez) will get the job, without at the same time acquiring any new belief instead. In all other respects the story is just the same as before. So this time, in Scenario $1'$, when Paul is given the information that Andrews or Becker might get the job, he withdraws his belief that $\neg a$, but he keeps $\neg b$. And in the alternative Scenario $2'$, when Paul learns from the dean that Andrews or Becker or Cortez might get the job, he again understands that competence in logic is regarded as an asset by the selection committee, and so he withdraws $\neg b$ while retaining $\neg a$. Scenario $3'$ just leads Paul to withdraw $\neg c$.

Now we can see that the prescriptions of the above principles for belief contraction are not complied with. First consider principle (5) and substitute $\neg a \wedge \neg b$ for x and $\neg c$ for y. Then we get $\neg c$ in $B \dot{-} (\neg a \wedge \neg b \wedge \neg c)$, but this belief set is not a subset of $B \dot{-} (\neg a \wedge \neg b)$, since $\neg a$ is in the former but not in the latter set.

Finally, the same substitutions serve to refute principle (6). The belief $\neg b$ is retained in both $B \dot{-} (\neg a \wedge \neg b)$ and $B \dot{-} \neg c$, but it is withdrawn in $B \dot{-} (\neg a \wedge \neg b \wedge \neg c)$.

In sum, then, we have found that Paul's reasoning which is perfectly rational and adequate for the situations sketched leads to belief formation processes that violate each of the six fundamental principles (1) – (6). How can this be explained?

4 Problems of rational choice are problems for belief formation

A first intuitive reaction to the puzzle is to simply deny that the example exhibits the formal structure that it has been represented as having here, and to claim

instead that the various messages we may receive from the dean are incompatible with one another. When the dean says that either Andrews or Becker or Cortez will be offered the job, isn't she, in some sense, saying *more* than when she says that either Andrews or Becker will be the winner of the competition? Namely, that it is *possible* that Cortez will be offered the position, while the latter message, at least implicitly, excludes that possibility. Shouldn't we therefore represent the dean's message in a somewhat more explicit way?

Three things can be said in reply to this objection. First, it is true that $a \vee b$ *implicitly* conveys the information that Cortez is not among the selected candidates. However, the kind of reasoning that turns implicit messages into explicit belief is exactly what is meant to be captured by theories of nonmonotonic reasoning and belief change. It is therefore important to insist that $\neg c$ is not part of the dean's message, but that it is rather inferred (perhaps subconsciously, automatically) by the reasoner. Representing the dean's statement in scenario 1 as $(a \vee b) \wedge \neg c$ would simply not be adequate.

Second, it is of course true that the message $a \vee b \vee c$ does not in itself exclude the possibility that c will come out true. But we must not think that each individual disjunct is considered to be a serious possibility by any of the interlocutors. For instance, nothing in the story commits us to the view that either the dean or Paul actually believes that Cortez stands a chance of being offered the position. So the dean's statement in scenario 2 must not be represented as saying that each of a, b and c is possible.

Finally there is a self-imposed limitation on expressiveness of the propositional language. As is common in the literature on belief formation, we presuppose in this paper that our language does not include the means to express autoepistemic possibility, something like $\Diamond c$ (read as, 'for all I believe, c is possible'). Admitting such means in the theory of belief formation immediately makes matters extremely complicated and invalidates almost all of the logical principles that have been envisaged for belief formation.[7]

We conclude that the problem if not caused by a sloppy translation of a commonsensical description of the case into regimented language.[8] What, then, does

[7]Some of the relevant problems are highlighted by Rott (1989) and Lindström and Rabinowicz (1999). — As for the underlying language, it is worth noting that our challenge to the theory of belief formation does not depend on any extension of standard propositional language, as other counterexamples to prominent logical principles do. Compare the much-debated riddles raised by McGee's (1985) counterexample to modus ponens and Gärdenfors' (1986) trivialization theorem, which both depend on the language's including non-truth functional conditionals.

[8]Still, it is hard to get rid of the feeling that the dean's information about the final candidates conveys more information than meets the ear. A logician gets mentioned as a top-ranking contender, and this alone has unexpected repercussions. Does this phenomenon point to an ordinary problem of belief formation, or does it constitute a problem *sui generis*? Does the very fact that something is the topic of a statement carry surplus meaning, over and above the latter's propositional content? My suspicion is that the fact that something is offered in a menu for inference, acceptance or removal has a special relevance that has so far been overlooked by

the problem arise from?

Principles of nonmonotonic inference and belief change can be systematically interpreted in terms of rational choice.[9] According to this view, the process of belief formation is one of resolving conflicts among one's beliefs and expectations by following through in thought the most plausible possiblities. According to a semantic modelling, the reasoner takes on as beliefs everything that is the case in all of the *most plausible models* that satisfy the given information, where the most plausible models are determined with the help of a selection function. A syntactic modelling, closely related to the semantic one, describes the reasoner as eliminating the *least plausible sentences* from a certain set of sentences that generates the conflict within his belief or expectation set. And again, the task of determining the least plausible sentences is taken over by a selection function. It is not possible here to give a description of these nicely dovetailing mechanisms even in the barest outlines. Suffice it to say that there are elaborate theories exhibiting in full mathematical detail striking parallels between the 'theoretical reason' at work in belief formation processes and those parts of 'practical reason' that manifest themselves in processes of rational choice.

On this interpretation, Disjunction in the Premises (2) and its counterparts for belief change, (4) and (6), turn out to be instantiations of one of the most fundamental conditions —perhaps *the* most fundamental condition—of the theory of rational choice. This condition, called *Independence of Irrelevant Alternatives*, the *Chernoff property* or *Sen's Property* α, says that any element which is optimal in a certain set remains an optimal element after some other elements are cancelled from that set. Cumulative Monotony (1) and its counterparts, (3) and (5), have turned out to be instantiations of another important condition in the theory of rational choice, namely to *Aizerman's axiom*.

The above scenarios are modelled after well-known choice situations in which Property α is violated, cases which also happen to disobey Aizerman's axiom. These properties may fail to be satisfied if the very 'menu' from which an agent is invited to choose carries information which is new to the agent.

The *locus classicus* for the problem is a passage in Luce and Raiffa (1957, p. 288). They tell a story about a customer of a restaurant who chooses salmon from a menu consisting of salmon and steak only, but changes to steak after being informed that fried snails and frog's legs are on the menu, too. This customer is not to be blamed for irrationality. The reason why he changes his mind is that he infers from the extended menu that the restaurant must be a good one, one where no risk is involved in taking the steak (which is the customer's 'real' preference as it were). Sen calls this phenomenon the 'epistemic value' or the

theorists of belief revision and defeasible reasoning.

[9]The theory of rational choice I am referring to here is the classical one deriving from economists like Paul Samuelson, Kenneth Arrow and Amartya Sen. A beautiful and concise summary of the most relevant ideas is given by by Moulin (1985). This theory is applied to the field of belief formation by Lindström (1991) and Rott (1993, 2001).

'epistemic relevance of the menu'.[10] Luce and Raiffa chose to avoid the problem by *fiat*:

> This illustrates the important assumption implicit in axiom 6 [= essentially Sen's Property α], namely, that adding new acts to a decision problem under uncertainty *does not alter one's a priori information as to which is the true state of nature*. In what follows, we shall suppose that this proviso is satisfied. In practice this means that, if a problem is first formulated so that the availability of certain acts influences the plausibility of certain states of nature, then it must be reformulated by redefining the states of nature so that the interaction is eliminated.

Luce and Raiffa thus explain away the problem of the restaurant customer because the extended menu conveys the information that the restaurant is a good one. The customer's choice is not really between salmon and steak, but basically between salmon-in-a-good-restaurant and steak-in-a-good-restaurant (assume that he does not like fried snakes and frog's legs).

Analogously, we may say that in the above example, Paul's doxastic choice is not simply one between the belief that Andrews gets the job and the belief that Becker gets the job. Given the information in scenario 2, his choice is rather between the belief that Andrews gets the job *and logic matters*, and the belief that Becker gets the job *and logic matters*. So it seems that the two scenarios cannot be compared in the first place. Have we now solved the puzzle?

No, we haven't. To see this, we have to understand first what is *not* responsible for the problem. In Luce and Raiffa's example, the reason for the trouble is not that the extended menu introduces a refinement in the customer's options, nor is it that his preferences change, nor is it that the second situation cannot be compared with the third. The customer is well aware right from the beginning that there are good restaurants and bad restaurants, and that he would prefer steak in a good, but salmon in a bad restaurant. What the availability of snails and frog's legs signals, however, is that the customer is actually in a good restaurant, whereas he had been assuming that he is in a bad one.[11] Luce and Raiffa

[10] Sen (1993, pp. 500–503; 1995, pp. 24–26) has brought the problem to wide attention. There are other reasons why Property α may fail without the chooser being irrational; see for example Levi (1986, pp. 32–34) and Kalai, Rubinstein and Spiegler (2002) about decision making on the basis of multiple preference relations. It remains to be seen how many of the reasons that speak against Property α as a general requirement for rational choice apply to the rather special domain of belief formation.

[11] Two sorts of reasons come to mind that might account for the customer's pessimism. Either his experience is that there are more bad restaurants than good ones, which makes is more likely that the one he is just visiting is a bad one. Or the pessimistic assumption is made because it is the relevant one for finding out which decision minimizes the customer's maximal damage, and the customer indeed wishes to be on the safe side.

are right in suggesting that the point is that the extended menu carries novel information about the state of the world.

Luce and Raiffa's argument thus may make good sense as a rejoinder in the context of the general theory of choice and decision. It is simply not this theory's business to explain how information is surreptitiously conveyed through the particular contents of a certain menu. So Luce and Raiffa have a justification for refusing to deal with that problem. Unfortunately, no analogous defense is available against the problem highlighted in the present paper. It *is* the business of theories of belief formation (which include expectation-based inference and belief change) to model how one's prior information is affected by information received from external sources. This is precisely what these theories have been devised to explain! Therefore, the anomaly cannot be pushed away into a neighbouring research field.

5 Conclusion

What is the moral of our story? We began by reviewing six of the most important and central logical principles that have generally been taken to be valid for commonsense reasoning and that have widely been endorsed as yardsticks for evaluating the adequacy of systems of non-classical logics intended to capture such reasoning. We have seen, however, that there are situations in which these reasoning patterns should not be expected to hold. This comes down to declaring them invalid, not as a contingent matter of some particular system that has been proposed in the literature, but as norms to which rational belief formation ought to conform.

The second lesson to be drawn from the above discussion is that a choice-theoretic modelling of belief formation processes does not only inherit the elegance and power of the theory of rational choice, but also its problems. This is not a trivial observation. Problems encountered in a general theory need not necessarily persist if this theory is applied to a restricted domain. The processes involved in belief formation are of a broadly logical kind, and one may perhaps expect that this domain is of a particularly well-behaved kind in which one would not encounter the strange phenomena surrounding the notion of rational choice. Our example has shown that this is not the case. The problems do carry over from the general to the more specific domain. Fundamental principles of belief formation are as affected by perturbations through the 'informational value of the menu' as the principles of rational choice.

Thirdly, we have found that things are even worse as regards this special domain. The reason is that a natural defense—Luce and Raiffa's defense—which makes sense for the general theory is not open for the special case of belief formation. The problem of the informational value of the menu may appear to be something which is alien to the concerns of rational choice theory. However, it

is clear that theories of belief formation are theories just about the processing of information that comes in propositional form. So the source of the trouble concerns a *paradigm problem* for belief formation theories rather than something that may be discharged into some other field of research.

So our discussion has a negative end. We have identified a big problem, but we haven't offered a solution for it. Once we make logic more 'realistic' in the sense that it captures patterns of everyday reasoning, there is no easy way of saving any of the properties that have endeared classical logic to students of the subject from Frege on. But problems there are, and creating awareness of problems is one of the important tasks of philosophy. It may be hoped that this paper instigates research that will eventually lead to more sophisticated models of belief formation.

Acknowledgement

I would like to thank Isaac Levi, David Makinson, Erik Olsson and Wlodek Rabinowicz for helpful discussions of earlier versions of this paper.

References

Alchourrón, Carlos, Peter Gärdenfors and David Makinson: 1985, 'On the Logic of Theory Change: Partial Meet Contraction Functions and Their Associated Revision Functions', *Journal of Symbolic Logic* **50**, 510–530.

Brewka, Gerd: 1991, 'Cumulative Default Logic: In Defense of Non-monotonic Inference Rules', *Artificial Intelligence* **50**, 183–205.

Dennett, Daniel: 1987, *The Intentional Stance*, MIT Press, Cambridge, Mass.

Gärdenfors, Peter: 1986, 'Belief Revisions and the Ramsey Test for Conditionals', *Philosophical Review* **95**, 81–93.

Gärdenfors, Peter: 1988, *Knowledge in Flux. Modeling the Dynamics of Epistemic States*, Bradford Books, MIT Press, Cambridge, Mass.

Gärdenfors, Peter, and Hans Rott: 1995, 'Belief Revision', in D. M. Gabbay, C. J. Hogger, and J. A. Robinson (eds.), *Handbook of Logic in Artificial Intelligence and Logic Programming Volume IV: Epistemic and Temporal Reasoning*, Oxford University Press, Oxford, pp. 35–132.

Giordano, Laura, and Alberto Martelli: 1994, 'On Cumulative Default Logics', *Artificial Intelligence* **66**, 161–179.

Hansson, Sven O.: 1999, *A Textbook of Belief Dynamics: Theory Change and Database Updating*, Dordrecht: Kluwer.

Kalai, Gil, Ariel Rubinstein and Ran Spiegler: 2002, 'Rationalizing Choice Functions by Multiple Rationales', *Econometrica* **70**, 2481–2488.

Kyburg, Henry E., Jr.: 2001, 'Real Logic is Nonmonotonic', *Minds and Machines* **11**, 577–595.

Levi, Isaac: 1986, *Hard Choices*, Cambridge University Press, Cambridge.

Lindström, Sten: 1991, 'A Semantic Approach to Nonmonotonic Reasoning: Inference Operations and Choice', Uppsala Prints and Preprints in Philosophy, Department of Philosophy, University of Uppsala, 1991:6.

Lindström, Sten, and Wlodek Rabinowicz: 1999, 'DDL Unlimited: Dynamic Doxastic Logic for Introspective Agents', *Erkenntnis* **50**, 353–385.

Luce, R. Duncan, and Howard Raiffa: 1957, *Games and Decisions*, Wiley, New York.

Makinson, David: 1994, 'General Patterns in Nonmonotonic Reasoning', in D. M. Gabbay, C. J. Hogger and J. A. Robinson (eds.), *Handbook of Logic in Artificial Intelligence and Logic Programming,* Vol. 3: *Nonmonotonic Reasoning and Uncertain Reasoning*, Oxford University Press, Oxford, pp. 35–110.

Makinson, David, and Peter Gärdenfors: 1991, 'Relations Between the Logic of Theory Change and Nonmonotonic Logic', in A. Fuhrmann and M. Morreau (eds.), *The Logic of Theory Change*, Springer LNAI **465**, Berlin etc., pp. 185–205.

McGee, Van: 1985, 'A Counterexample to Modus Ponens', *Journal of Philosophy* **82**, 462–471.

Morgan, Charles G.: 2000, 'The Nature of Nonmonotonic Reasoning', *Minds and Machines* **10**, 321–360.

Moulin, Hervé: 1985, 'Choice Functions over a Finite Set: A Summary', *Social Choice and Welfare* **2**, 147–160.

Reiter, Raymond: 1980, 'A Logic of Default Reasoning', *Artificial Intelligence* **13**, 81–132.

Roos, Nico: 1998, 'Reasoning by Cases in Default Logic', *Artificial Intelligence* **99**, 165–183.

Rott, Hans: 1989, 'Conditionals and Theory Change: Revisions, Expansions, and Additions', *Synthese* **81**, 91–113.

Rott, Hans: 1993, 'Belief Contraction in the Context of the General Theory of Rational Choice', *Journal of Symbolic Logic* **58**, 1426–1450.

Rott, Hans: 2001, *Change, Choice and Inference*, Oxford Logic Guides, Vol. 42, Clarendon Press, Oxford.

Sen, Amartya K.: 1993, 'Internal Consistency of Choice', *Econometrica* **61**, 495–521.

Sen, Amartya K.: 1995, 'Is the Idea of Purely Internal Consistency of Choice Bizarre?', in J. E. J. Altham and R. Harrison (eds.), *World, Mind, and Ethics. Essays on the Ethical Philosophy of Bernard Williams*, Cambridge University Press, Cambridge, pp. 19–31.

Stalnaker, Robert C.: 1984, *Inquiry*, Bradford Books, MIT Press, Cambridge, Mass.