

REASONING FROM INCONSISTENCY: A FIELD EXPLORATION

Kristien DIEUSSAERT & Walter SCHAEKEN
University of Leuven

Preface

This issue has proceeded from the Fifth International Conference on Thinking, which was held in the Department of Psychology of the University of Leuven, Belgium, July 22-24, 2004. The conference, which was sponsored by the Fund for Scientific Research Flanders and the British Psychological Society (Cognitive Section), intended to bring together researchers working in different domains of the psychology of thinking and reasoning. Five keynote speakers (Ruth Byrne, Vinod Goel, Karl-Christoph Klauer, Paolo Legrenzi, & Douglas Medin) were introducing some of the important topics (the rational imagination, the neuropsychology of reasoning, the Wason selection task, reasoning to consistency, biological thought within and across cultures).

We want to use this issue to direct the reader's attention to trends in the psychology of reasoning from inconsistency, which is a recent and growing topic. We do have, however, a second reason for choosing this topic as the main theme of this special issue. Indeed, this issue enables us to honour the vital role of Paolo Legrenzi in improving the quality and quantity of the European research in the psychology of reasoning. Paolo Legrenzi has not only published numerous fundamental articles in the most important journals and chapters in essential books, he also inspired and motivated many other European researchers. With his pleasurable character and bright mind he could easily stimulate people to start and to keep working in the domain of the psychology of reasoning. We can only be very grateful for his work and personality.

We hope to show in this issue a sample of the diversity of research related to reasoning to consistency, the topic of Paolo Legrenzi's most recent research and a topic that has drawn the attention of researchers from various disciplines, such as psychology, philosophy, computer science and so on. Although the work presented is mainly psychological, the reader will discover that it is heavily inspired by the aforementioned disciplines.

Laboratory of Experimental Psychology, Psychology Department, University of Leuven.
This paper was written with support from FWO G.0239.02.

Correspondence: Kristien Dieussaert, Laboratory of Experimental Psychology,
Tiensestraat 102, 3000 Leuven. E-mail: Kristien.Dieussaert@psy.kuleuven.ac.be

Two main streams are found in the literature on reasoning from inconsistency: belief revision and nonmonotonic reasoning. Belief revision is temporally oriented. It deals with creating consistency in a set of beliefs which has become inconsistent through the introduction of new information. Nonmonotonic reasoning focuses on the determination of the inferences that are allowed to be drawn from an inconsistent set of premises. In the remainder of this introductory paper, we elaborate more on experimental research in both streams. Also in this issue, both streams of research are represented: the first three papers deal with belief revision, while the latter two focus on nonmonotonic reasoning.

New information inconsistent with the present belief structure generally triggers a process of belief revision. In some cases, one may come up with diagnostic explanations for the inconsistency. However, one diagnostic explanation is not the other. In Paolo Legrenzi and Philip N. Johnson-Laird's paper, we read about preferred explanations to solve the inconsistency. In other cases, one may rather disregard some of his or her previous beliefs. Various paradigms have been introduced in the literature to study belief revision. In a methodological paper, Kristien Dieussaert, Wim De Neys and Walter Schaeken discuss three important paradigms to reasoning from inconsistency. Russell Revlin, Dustin Calvillo, and Stephanie Ballard introduce a new paradigm, in which belief revision is measured through the observation of actual behaviour rather than through verbal reports.

Marilyn Ford focuses on the -often not rational- factors that people take into account when drawing inferences from inconsistent information. These factors do not seem to correspond with the intuitive ideas on human inference making that are often taken for granted in nonmonotonic formalisations. In the same line of thinking, Niki Pfeifer and Gernot Kleiter propose a probabilistic nonmonotonic formalisation that is not only inspired by logic tradition, but relies heavily on psychological data. The crossfertilization between logic and psychology, also present in the other papers of this issue, leads towards a better understanding of common sense reasoning, i.e., reasoning in a complex world, where inconsistencies are omnipresent.

It is our sincere hope that this special issue will inspire cognitive scientists to start or proceed with research in this very rich and promising domain.

An introduction to reasoning from inconsistency

In this section we give an introduction to the theories and research on reasoning from inconsistency. We mainly take psychological theories into account, but it will become clear throughout the section that psychological research on this topic is often inspired by other disciplines, such as philoso-

phy and AI. The section is split up into two main parts. In the first part we report on (psychological) theory development and in the second part we review experimental research on reasoning from inconsistencies. We focus on research involving conflict (or semi-conflict) but do not broaden this review to related research that is nonmonotonic as well, such as statistical reasoning (e.g., Girotto & Gonzalez, 2001), diagnostic reasoning (e.g., Kim & Keil, 2003), inductive reasoning (e.g., Heit, 2000), or other related domains.

Theoretical developments

Over the years, the theories developed to explain human reasoning have mainly focused on deductive tasks (propositional arguments, syllogisms) and on hypothesis testing (Wason selection task). Within this context it was not necessary for these theories to explain how people deal with inconsistent information. However, since individuals are prepared to reject former inferences and since daily life requires them to deal with information inconsistent with their former beliefs, one may expect from any descriptive reasoning theory that it includes defeasibility in its framework. Below, we briefly describe four theories of human reasoning, and how they accommodate for reasoning from inconsistency.

Natural deduction

Many older psychological theories of reasoning postulate that people are equipped with formal rules of inference akin to those of a logical calculus (see e.g., Braine, Reiser, & Rumin, 1984; Osherson, 1975; Rips, 1983, 1994). We will refer to them in the following as natural deduction theories. These theories are syntactic theories: they claim that deductive reasoning consists of the application of inference rules to the form of the premises and conclusion of an argument. People have little difficulty in extracting the logical form from a conditional utterance, but they only have a limited set of formal rules available. According to natural deduction theories, our mind contains a rule for Modus Ponens.

If p , then q . $p \rightarrow q$

Consider a Modus Ponens problem:

If there is a circle on the blackboard, then there is a square on the blackboard.

There is a circle on the blackboard.

The rule or reasoning schema matches the form of the problem. Therefore, the inference can be made promptly: there is a square on the blackboard.

Consider, however, a Modus Tollens problem:

If there is a circle on the blackboard, then there is a square on the

blackboard.

There is no square on the blackboard.

The correct solution to this Modus Tollens problem is: there is no circle on the blackboard. According to natural deduction theories, there is no rule in our mind corresponding to this kind of problem. Therefore, it is only indirectly – through several reasoning steps – that we can come up with the correct solution. According to the natural deduction theories, this explains why solving Modus Tollens problems takes longer and is more prone to errors than solving Modus Ponens problems.

Rips (1994) is an advocate of the natural deduction account. In ‘The psychology of proof’, in which he expounds the PSYCOP model of human reasoning, he also dedicates attention to nonmonotonic reasoning. He argues that his deduction system can deal with default rules, because: (1) there is no need to suppose that every belief in the database either is an axiom or follows deductively from axioms (i.e., new information might stem from induction, or a default rule, which is translated into ‘a plausible or inductively strong conclusion from our general knowledge) and (2) as the Truth Maintenance System (TMS; McDermott & Doyle, 1980) does, PSYCOP is able to detect inconsistencies in a database because it records all the dependency-relations (the justifications on which a belief is based). Rips admits that PSYCOP should to this end be extended to also keep track of the information that is put in the database in an inductive manner (Harman, 1986). Moreover, there is no mechanism (in TMS nor PSYCOP) that dictates which beliefs should be revised and which not (only the assumptions and premises on which the conclusion is based are mentioned) while this is supposed to be a crucial property of a system that tries to describe nonmonotonic reasoning. Revlin, Cate and Rouss (2001) put forward a proposal in this sense for natural deduction but this needs further study.

Rips (1994) recognizes the (temporary) failure of purely deductive systems (such as PSYCOP) to explain default reasoning. His main point, however, is that nonmonotonic logics (e.g., circumscription) are less than ideal for cognitive purposes because they do not lend themselves to simple implementations and because they do not seem to reflect the deliberations that actually underlay human reasoning with defaults.

Generality coding model

With the new interest in common sense reasoning, an older theory of Revlis (1974; see also Revlin et al., 2001) on counterfactual reasoning has regained interest because of its applicability to belief revision. Counterfactual reasoning is reasoning from facts that are assumed not to be true (e.g., if the train had been in time, I would have missed it). The generality coding model (GCM) of Revlis (1974) is based on the principles of modal logic.

Belief contravening problems (Rescher, 1964) are at the heart of this theory. In belief contravening problems, the reasoner is given statements of the form: (1) All X are Y, (2) This Z is not a Y, (3) This Z is not a X, *(4) This Z is a X. The reasoner is asked to assume that the statement with the asterisk (4) is true and to reconcile this information with previous beliefs (1, 2 or 3). It will be clear that giving up (3), implies that the reasoner also has to give up statement (1) or (2) to regain a consistent belief set. Revlin et al. (2001) call these kinds of problems 'combining problems' because something is added to the set of beliefs. The structure of these problems resembles the Modus Tollens problem with a conflicting conclusion (4). Rending problems, in which something is retracted from the set of beliefs, have a structure that resembles the Modus Ponens problem with a conflicting conclusion (4): (1) All X are Y, (2) This Z is a X, (3) This Z is a Y, *(4) This Z is not a Y.

The GCM states that reasoning with these problems proceeds in three stages. First, the reasoner constructs a possible world in which the counterfactual assumption (4) can be true. Second, the reasoner orders the relevant, available descriptions about the world in terms of his/her ability to structure the domain, in this way creating an entrenchment ordering in terms of modal categories. An example is differentiating among statements that are necessarily true (e.g., laws) and those that are contingently true (e.g., particular statements). Third, the reasoner seeks to resolve inconsistencies between statements by retaining the one with the lowest modal status (i.e., the most necessary proposition).

For the combining problem, if statement (1) expresses a lawlike relation, it will be the more entrenched one and the particular sentence (2) will be revised. This is so because the more general a proposition is, the more instances it predicts and thus the more central it is for reasoning about new, possible worlds. The combination of the assumption (4) and the generality (1) leads to a revision of (2). However, if the relation in (1) is an 'accidental' one, no preference relation can be set, and thus both statements (1) or (2) are prone to be chosen for revision. Indeed, besides the choice for the generalist strategy (combining 4 and 1), the reasoner may now adopt a particularist strategy (combining 4 and 2) since both 1 and 2 are equally entrenched.

For the rending problem, the generality (1) has the highest degree of necessity to begin with. The particular statement (3) has lost only its distant category membership by virtue of the counterfactual assumption (4). It still possesses its original, immediate category membership (2). In the case of a combining problem, under the assumption (4, e.g., this whale animal is a mammal), a possible world could be constructed that immediately showed the predictive power of the generality. In this case, the generality has lost this special status. If (4) is assumed (e.g., this animal is not a mammal), the generality (1; e.g., all whales are mammals) only has the status of category mem-

bership, but its central status as a predictive element is gone. Thus, the contrast is between two immediate category memberships: the generality (1), which assigns all members of one category into another superordinate category (e.g., whales → mammals), and the particular statement (2), which maintains the assignment of an instance to an immediately superordinate category (e.g., this animal → whale). From a modal perspective, none of them is as such more entrenched than the other statement. Both statements are thus prone to revision.

To summarize, the GCM makes predictions about the belief revision process based on a modal logic perspective. Although it was developed as a model with a very small scope, viz. to explain how belief contravening problems were solved, it turns out to be valuable for other kinds of problems that involve a sort of conflict too. See also Revlin, Calvillo, and Ballard (2005) further in this issue.

Mental model theory

According to the mental model theory (MMT; Johnson-Laird, 1983; Johnson-Laird & Byrne 2002), reasoning consists of three main stages. First, the premises are understood: a mental model of the situation they describe is constructed on the basis of their meaning and of any relevant general knowledge triggered during the process of interpretation. Each model represents a true possibility. Second, reasoners formulate a conclusion based on the model. People will only draw conclusions that convey some information that was not explicitly asserted by the premises. Third, a search is made for alternative models of the premises in which the putative conclusion is false. If there is no such model, then the conclusion is valid; that is, it must be true given that the premises are true. If there is such a model, then it is necessary to return to the second stage to determine whether there is any conclusion that holds over all the models constructed so far. The theory's essential processing assumption is that the more models that have to be constructed, the harder the inferential task will be. Nonmonotonic processes belong to the core of MMT, since it allows arbitrary or default components in a model that can be corrected when more information is given (e.g., Johnson-Laird, 1993).

However, in case of inconsistent models this explanation does not suffice. To fill this gap, Girotto, Johnson-Laird, Legrenzi and Sonino (2000; Johnson-Laird, Girotto, & Legrenzi, 2004) have worked out some principles that, added to the MMT, result in a theory of *naïve* nonmonotonic reasoning (nNMR).

Again, three stages are discerned within the nNMR theory. During the first stage, reasoners evaluate whether the mental models they built are internally consistent (the principle of modeling consistency). Reasoners try to fit all the propositions in a consistent model, by giving up arbitrary or default assump-

tions. If the propositions cannot be fitted in such a model, they infer that the propositions are inconsistent.

For the second stage, two possible scenarios are described when the reasoner decides that a set of propositions is inconsistent. A first scenario may occur if the reasoner knows no proposition with a greater certainty than the others: the reasoner may choose to defer any attempt to reach consistency until (s)he has more information. A second scenario may occur if the inconsistency arises from a conflict between incontrovertible evidence and a valid inference from propositions: whichever proposition mismatches the evidence is rejected (the mismatch principle). This mismatch principle can be applied differently, depending on the situation at hand. If a proposition directly conflicts with the evidence, the proposition will be rejected. If the propositions are not equiprobable, the most improbable proposition is rejected. In all other cases, the proposition that has to be rejected is the one for which only mental models conflicting with the evidence or failing to represent the evidence exist. Imagine a reasoner being presented with a Modus Ponens problem 'If there is a circle on the blackboard, then there is a square on the blackboard' and 'There is a circle on the blackboard'. (S)he knows for sure that 'There is no square on the blackboard'. The evidence (no square) is consistent with each of the previous assertions separately, but the propositions cannot be fitted all three in one model. Thus, one proposition should be rejected. The evidence conflicts with the initial model of the conditional, so this is the one that should be rejected. However, the reasoner could have represented the conditional premise fully explicit, as would have been necessary to solve the Modus Tollens problem. This would have led to the rejection of the categorical premise because its model fails to represent the evidence while one of the conditional premise's models matches the evidence. (see Johnson-Laird et al., 2004 for a visualisation of the mental models).

During the third stage, this is after the detection of inconsistency and rejection of the mismatching proposition, reasoners use their available causal knowledge to create explanations of what led to inconsistency (the principle of causal knowledge). This knowledge is represented in explicit models which can modulate the mental models of assertions, taking precedence over them in the case of contradictions. Further in this issue, Legrenzi and Johnson-Laird (2005) go deeper into the role and preference of diagnostic explanations to inconsistencies.

A probabilistic account

A more recent theory on human reasoning, put forward by Oaksford and Chater (e.g., 1998), gives a probabilistic account of reasoning. Their proposal is based on a rational analysis (see Anderson, 1991) of the Wason selection task (Wason, 1966). Oaksford and Chater argue that most of the results

that are over the years obtained with this task can be explained by assuming that people hold in mind subjective probabilities.

In Chater and Oaksford (1999) they give a very clear description of their ideas in terms of rational analysis copied here. In the Wason selection task, the participant's goal can be clearly defined as selecting the card that contains the greatest expected informativeness about whether the rule 'if p , then q ' is true or not (given, p and q are independent). Oaksford and Chater (1998) apply an Optimal Data Selection (ODS) model (based on Bayesian statistics) that relies on the crucial assumption that properties are rare (i.e., most of the time, no circles or squares appear on a blackboard). Under this assumption, it makes sense to say that the most informative 'experiments' to test the truth of the rule 'if there is a circle on the blackboard, then there is a square on the blackboard' are to watch for a square if a circle is present (p -card), which is one of the 'correct answers'. Moreover it makes sense to see whether there is a circle, if a square is present (q -card), but it does not make sense to do what is generally considered as the 'correct answer': turning the not- q card (i.e., no square). The expected information gain confirms this intuition. The gain is higher for p than for q , which in its turn has a higher expected gain than not q . The expected gain is lowest for not p . The number of cards that the participant in an experiment will turn over depends on a cost (turning cards) – benefit (gaining information) trade-off. Oaksford and Chater (1998) show that the ODS model can account for most of the data on the Wason selection task. Their account is also applicable to conditional reasoning in general (see Oaksford, Chater, & Larkin, 2000; and for a critique, see Schroyens & Schaeken, 2003).

In fact, their main argument is that reasoning research should not be studied in relation to classical logic, as the former three classes of theories propose, but that it should be viewed from a different perspective: that of the uncertainty in the world with which individuals have to deal daily.

With respect to nonmonotonic reasoning, Oaksford and Chater claim that probabilities alone could never offer a (complete*) theory of everyday reasoning, since they observed that everyday reasoning cannot so easily be formalised, even not within probabilistic frameworks. The crucial issue according to them is to look at dependencies between propositions (How does learning A change the probability of C if given B ?). With Pearl (1988), they postulate that these questions are fundamentally qualitative in character, and are thus more important than numerical calculations. Pfeifer and Kleiter (2005) extend further on it later in this issue.

The four theories presented all dedicate attention to reasoning from inconsistencies. They each rely on very different explanatory mechanisms, viz., mental inference rules, modal properties, mental models and subjective prob-

abilities. It seems difficult to assume that these theories will once be reconciled in one comprehensive theory of common sense reasoning. In the end, experimental research will decide which of the ideas brought forward in the four theories will survive the test.

Experimental reasoning research

In this section we will review reasoning research that involves dealing with (semi-)conflictuous information. Although the amount of research on this topic is fairly limited, four different categories of research can already be distinguished. A first category is research in which no direct conflictuous information is present, but in which plausible conflicting evidence is made salient. The experiments test which information is of special relevance to suppress valid inferences. This is the 'oldest' category and therefore involves more experiments than the other two categories. A second category is research that focuses on belief revision. These experiments test which propositions individuals are more inclined to revise than others, when given conflictuous information. A third and fourth category represent research with the focus on nonmonotonicity. The experimental studies of the third category test whether the specific properties of formal nonmonotonic systems also hold for human reasoning, while the studies of the fourth category put the benchmark problems to the test.

Reasoning in semi-conflict

The 'suppression' research line in reasoning started off with Byrne's paper (1989), in response to a study of Rumin, Connell, and Braine (1983). She presented participants with statements such as the following:

If it rains, then you will get wet.

If you walk outside, then you will get wet.

It rains.

Byrne (1989) showed that the Modus Ponens inference was suppressed by adding a logically irrelevant, but semantically relevant, premise to the classical Modus Ponens-problems (MP; if p , then q , p).

The semantic relevance of this premise lies in the fact that it triggers a possible exception to the conditional premise (namely staying inside) that could be formalized as follows:

If p (rain) then q (wet).

If $p \& r$ (inside and rain), then not- q (not-wet).

Given p , participants become uncertain whether the more specific rule applies here or not, resulting in fewer persons accepting the MP conclusion 'q'. Thus, although there is no direct conflict apparent and one could easily deny the second conditional premise, the additional information is taken into

account (in accordance with Sperber & Wilson, 1986 and Grice's maxim of relation) and creates an inconsistent state that participants wish to resolve. Further in this issue, Dieussaert, De Neys, and Schaeken (2005) elaborate on this.

Belief revision

Research on belief revision has only very recently become a topic of interest within reasoning research. In this line of research, participants are given a conditional statement (if p , then q) or a universal quantifier (all p are q) and a categorical statement (e.g., p), and are asked to deduce the conclusion, or are given the conclusion (e.g., q). Next, new information that contradicts the conclusion is given (e.g., not- q) and participants are asked to revise one of the former statements in order to get a consistent belief set again. Further in this issue, Legrenzi and Johnson-Laird (2005), Dieussaert et al. (2005) and Revlin et al. (2005) conduct their studies on belief revision with this paradigm.

Elio and Pelletier (1997; see also Elio & Pelletier, 1994) showed that the conditional premise rather than the categorical premise, is revised when an inconsistency arises. Since that pioneer study, two main findings have come to light. The first one is that the initial belief in the conditional plays an important role: the lower the initial belief in the conditional rule, the more revision of the conditional rule takes place (see Dieussaert, De Neys, Schaeken, & d'Ydewalle, 2000; Politzer & Carles, 2001; Calvillo & Revlin, 2002; Verhaert, 2004).

The second one is that the belief revision choice depends on the logical structure in which the inconsistency is introduced. This effect is dubbed the Inference Contradiction effect by Byrne and Walsh (2002). The general tendency is that the conditional is more revised for MP than for MT, albeit that the amount of revisions differs widely across different studies (from 3% to 84%; Elio & Pelletier, 1997; Dieussaert et al., 2000; Revlin et al., 2001; Calvillo & Revlin, 2002; Hasson & Johnson-Laird, 2002; Revlin & Calvillo, 2005). In Politzer and Carles (2001), Byrne and Walsh (2002), and to a smaller extent Revlin and Calvillo (2005, Exp. 2), the Inference Contradiction effect goes in the opposite direction compared to the other studies. So far, no theoretical explanation for this finding exists, although Verhaert (2004) has made the interesting observation in two experiments that the conditional is more revised in MP problems than in MT problems for 'diagnostic conditionals', while the reverse is true for 'definitional conditionals'. This recent finding may shed a new light on the Inference Contradiction effect.

To conclude this part, we would like to mention that only two (important) findings of the whole set of studies conducted on belief revision up till now

have been highlighted. Other findings such as order effects (e.g., Revlin & Calvillo, 2005), effects of negations in the conditional (Calvillo & Revlin, 2002), counterfactual conditionals (Byrne & Walsh, 2002) and so on were omitted here. For reasons of simplicity, we also refrained from mentioning studies using another format (e.g., Elio, 1997; Dieussaert, Schaeken, & d'Ydewalle, 2002a, 2002b), although they also provide interesting information on the topic.

Testing nonmonotonic properties: the rationality postulates

Elio and Pelletier's study (1997) has its roots in AI logic (e.g., Gärdenfors & Rott, 1995). The subsequent experimental studies, however, paid far less attention to the logical aspects as well as to the consequences of their findings for logical formalisations.

The experimental study of rationality postulates has its roots also in AI logic. In this case however, the main aim of the studies is to strengthen or weaken logical formalisms with the aid of psychological data. AI logicians have developed various sets of meta-theoretical properties that nonmonotonic inference relations should obey to. Kraus, Lehmann, and Magidor (1990) describe five such metatheories of desirable consequence relations in a canonical paper. They dubbed the five systems as follows: C ('cumulative'), Cm ('C+monotony'), Cl ('C+loop'), P ('preferential') and M ('monotonic'). For a discussion of all these systems, see Kraus et al. (1990) and Brewka, Dix, and Konolidge (1997). Of all these systems, system P is generally considered the most appealing, as Kraus et al. (1990, p. 204) state:

Of those families of consequence relations, which is the best suited to represent the inferences of a nonmonotonic reasoner in the presence of a fixed knowledge base? Monotonic and cumulative monotonic reasoning are too powerful, i.e. simple cumulative and simple preferential models are too restrictive to represent the wealth of nonmonotonic inference procedures we would like to consider. We feel that all bona fide logical systems should implement reasoning patterns that fall inside the framework of cumulative reasoning, but probably not all cumulative models represent useful nonmonotonic systems. The same may probably be said about cumulative ordered models. Preferential reasoning seems to be closest to what we are looking for.

A lot of nonmonotonic logics share the rationality postulates of system P.

This has inspired several experimental psychologists to test (some of) the rationality postulates of P. Da Silva Neves, Bonnefon, and Raufaste (2002) tested these properties plus the Rational Monotony property with thematic material. They consider these properties as rationality postulates (thus as a norm of rationality) but not as descriptive for how human reasoning actually works. They argue that these properties should not be studied as is done tra-

ditionally (i.e. presenting a conditional and a categorical premise, and asking which conclusion follows necessarily from both premises) since they are not considered as direct inference rules. In a pilot study, Da Silva Neves et al. (2002) tested whether the relations they presented were interpreted as plausible or non-plausible rules, or as a material implication or a material equivalence.

The items were then filled corresponding to the syntax of the properties, and it was tested whether the same participants also accepted the 'conclusion' (Right Part in their terminology). For instance, the acceptance of following items was tested (in random order) for Rational Monotony (RM):

Lawyers have a large income (plausible rule, resulting from pilot study).

Lawyers do not speak Italian (implausible rule, resulting from pilot study).

Lawyers who speak Italian have a large income.

This corresponds with the syntax:

$a \mid \sim g, \ddot{y}(a \mid \sim \ddot{y} b) \text{ fi } a\ddot{y}b \mid \sim g$

Their results suggest that individuals accept most of the properties of system P, but that RM was not accepted. Additionally, it was found that the participants did not accept the Monotony property, albeit that they corroborated Cautious Monotony: $a \mid \sim b, a \mid \sim g \text{ fi } a\ddot{y}b \mid \sim g$

Pfeifer and Kleiter (in press) criticized on the methodology used by Da Silva Neves et al. (2002). Pfeifer and Kleiter regard the properties of P as proper inference rules, and argue that they should be tested as such. This was done successfully by Benferhat, Bonnefon and Da Silva Neves (in press). They found similar results although only 8% answered all questions according to the rationality postulates of system P. In this experiment 47% of the participants also accepted the Monotony property, which is classically valid but not P-valid.

Besides the aforementioned critique, Pfeifer and Kleiter (in press) pointed out another methodological flaw in the experiment of Da Silva Neves et al. (2002). They argue against the use of words such as 'typically, generally, usually' to express the idea of a default inference rule. They also argue against dropping the indication of nonmonotony (birds fly). Instead, they plead for the use of percentages or other specific indications of nonmonotony (e.g., 90% of the birds fly) and tested several rationality postulates for which they all found corroboration to a great extent. For more on rationality postulates, see Pfeifer and Kleiter (2005) further in this issue.

The Benchmark problems: defaults and specificity

As is shown in the previous paragraph, AI researchers have different views upon what 'the' rational answer is to a problem that involves inconsistency. Lifschitz (1988), among others, recognized this and set out a list of

inferences that are supposed to be valid in any proposed default reasoning system. This list of benchmark problems should grow and change over time, as he saw it. The idea behind the list is that any formalization of default reasoning should follow the benchmarks in telling which element should over-rule the other. He stated:

Ideally, we would like to have a single system of nonmonotonic reasoning that leads to correct and concise solutions for all the benchmark problems. But formalisms with limited possibilities can be valuable too [...] (Lifschitz, 1988, p. 203).

However, currently the problem is that often formalizations are developed only *in order to* solve some of these benchmark problems, rather than simply being able to solve these problems. This makes these systems questionable with respect to their general usefulness as a norm for rational human reasoning.

Another problem with these benchmark problems is that their solutions are merely based on the intuitions of a small group of AI researchers, while they are claimed to be ‘common sense solutions’. Several researchers familiar with experimental research in psychology, tested these AI intuitions on their generalisability. Ford and Billington (2000; see also Ford, *in press*) focussed on ‘specificity’, a property that rational agents require to solve the benchmark problems correctly. They presented participants unfamiliar material with different inheritance relations such as:

Hitta’s are usually not waffs.
(*i.e. a non-strict, default inheritance relation*)

All of the hitta’s are oxers.
(*i.e. a strict relation*)

Oxers are usually waffs.

Jukk is a hitta.

Is jukk a waff?

Penguins do not fly.

All penguins are birds.

Birds fly.

Tweety is a penguin.

Does Tweety fly?

This example is a translation of the more familiar Tweety triangle (e.g., Lifschitz, 1988). Ford and Billington (2000) argue against the use of familiar problems because they can be solved without reasoning, relying on background knowledge. They showed that for less familiar versions of the Tweety triangle, and other, intuitively categorized as simple nonmonotonic problems, participants have a lot of difficulties to solve them. Ford and Billington (2000) therefore claim that nonmonotonic reasoning is hard to do, in the same vein as it is hard for humans to reason correctly according to classical logic (see also Vogel, 1996; Hewson & Vogel, 1994). Participants in their

study also seemed reluctant to use ‘specificity’ as defined in the AI literature to solve the problems (For an alternative view, see Schurz, 2001). However, Ford and Billington believe that under the right circumstances humans can reason correctly nonmonotonically (e.g., graphical presentations might help). For more on the (absent) ability of human agents to reason nonmonotonically, see Ford (2005) further in this issue.

Pelletier and Elio (in press; see also Elio & Pelletier, 1993) tested Lifschitz’ basic default reasoning benchmark problems and inheritance benchmark problems. They obtained rather high correlations between the participant’s answers and the AI answers. Nevertheless, their results also showed that other properties, which are generally not taken into account by AI researchers, such as perceived similarity, amount of information etc., play a role in the answers the participants gave to the problems.

From these results Pelletier and Elio draw a more extreme conclusion than Ford and Billington (2000) did. Pelletier and Elio (1997) plead for a psychological interpretation of default reasoning: good default reasoning is only discovered by looking at how people behave, not by looking at mathematical systems or computational considerations.

To conclude on the nonmonotonic properties, some data confirm the intuitions of AI researchers and others disconfirm them. Considering the very recency of the studies and the small amount of research on this topic thus far, we believe that the only conclusion that can be drawn for the time being is that of Da Silva Neves et al. (2002, p. 128).

These preliminary results encourage us to engage in the search for new evidence using other materials, with the same experimental device as well as with new ones. We hope that these new studies will lead to conclusive data ... Finally, we believe that, in the long run, this new line of research is of interest for both the psychology and artificial intelligence communities.

References

- Anderson, J. R. (1991). The adaptive nature of human categorization. *Psychological Review*, 98 (3), 409-429.
- Benferhat, S., Bonnefon, J.-F., & Da Silva Neves, R. (in press). An overview of possibilistic handling of default reasoning, with experimental studies. *Synthese*.
- Byrne, R. M. J. (1989). Suppressing valid inferences with conditionals. *Cognition*, 31, 61-83.
- Byrne, R. M. J. & Walsh, C. (2002). Contradictions and counterfactuals: Generating belief revisions in conditional Inference. In W. D. Gray & C. D. Schunn (Eds.), *Proceedings of the 24th Annual Conference of the Cognitive Science Society* (pp. 160-165). Mahwah, NJ: Erlbaum.

- Braine, M. D. S., Reiser, B. J., & Rumin, B. (1984). Some empirical justification for a theory of natural propositional logic. In G. Bower (Ed.), *The psychology of learning and motivation, Volume 18*. (pp. 313-371). New York: Academic Press.
- Brewka, G., Dix, J., & Konolidge, K. (1997). *Nonmonotonic reasoning: An overview*. Stanford: CSLI Lecture Notes 73.
- Calvillo, D. & Revlin, R. (2002). The role of logical structure and premise believability in belief revision. In W. D. Gray & C. D. Schunn (Eds.), *Proceedings of the 24th Annual Conference of the Cognitive Science Society* (p. 993). Mahwah, NJ: Erlbaum.
- Chater, N. & Oaksford, M. (1999). Ten years of the rational analysis of cognition. *Trends in Cognitive Sciences*, 3 (2), 57-64.
- Da Silva Neves, R. M., Bonnefon, J. F., & Raufauste, E. (2002). An empirical test of patterns for nonmonotonic inference. *Annals of Mathematics and Artificial Intelligence*, 34, 107-130.
- Dieussaert, K., De Neys, W., & Schaeken, W. (2005). Suppression and belief revision, two sides of the same coin? *Psychologica Belgica*, 45 (1), 29-46.
- Dieussaert, K., Schaeken, W., De Neys, W., & d'Ydewalle, G. (2000). Initial belief state as predictor of belief revision. *Current Psychology of Cognition*, 19 (3), 277-288.
- Dieussaert, K., Schaeken, W., & d'Ydewalle, G. (2002a). The relative contribution of content and context factors on the interpretation of conditionals. *Experimental Psychology*, 49 (3), 181-195.
- Dieussaert, K., Schaeken, W., & d'Ydewalle, G. (2002b). The impact of the nature of disabling conditions on the reasoning process. *Current Psychology Letters*, 7, 87-103.
- Dieussaert, K., Schaeken, W., Schroyens W., & d'Ydewalle, G. (2000). Strategies during complex conditional inferences. *Thinking and Reasoning*, 6 (2), 125-160.
- Elio, R. (1997). What to believe when inferences are contradicted. The impact of knowledge type and inference rule. In M. G. Shafto & P. Langley (Eds.), *Proceedings of the 19th Annual Conference of the Cognitive Science Society* (pp. 211-216). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Elio, R. & Pelletier, F. J. (1993). Human benchmarks on AI's benchmarks problems. In *Proceedings of the 15th Annual Conference of the Cognitive Science Society* (pp. 406-411). Lawrence Erlbaum, Hillsdale, NJ.
- Elio, R. & Pelletier, F. J. (1994). The effect of syntactic form on simple belief revisions and updates. In A. Ram & K. Eiselt (Eds.), *Proceedings of the Sixteenth Annual Conference of the Cognitive Science Society* (pp. 260-266). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Elio, R. & Pelletier, F. J. (1997). Belief change as propositional update. *Cognitive Science*, 21 (4), 419-460.
- Ford, M. & Billington, D. (2000). Strategies in human nonmonotonic reasoning. *Computational Intelligence*, 16, 446-468.
- Ford, M. (in press). Human nonmonotonic reasoning: the importance of seeing the logical strength of arguments. *Synthese*.
- Ford, M. (2005). On using human nonmonotonic reasoning to inform artificial systems. *Psychologica Belgica*, 45 (1), 57-70.

- Gärdenfors, P. & Rott, H. (1995). Belief revision. In D. M. Gabbay et al. (Eds.), *Handbook of logic in artificial intelligence and logic programming*. Vol. 4: *Epistemic and temoral logics* (pp. 35-132). Oxford: Clarendon Press.
- Giroto, V. & Gonzalez, M. (2001). Solving probabilistic and statistical problems: A matter of information structure and question form. *Cognition*, 78 (3), 247-276.
- Giroto, V., Johnson-Laird, P. N., Legrenzi, P., & Sonino M. (2000). Reasoning to consistency: How people resolve logical inconsistencies. In J. A. Garcia-Madruga et al. (Eds.), *Mental models in reasoning* (pp. 83-98). Madrid: UNED Ediciones.
- Harman, G. (1986). *Change in view: Principles of reasoning*. Cambridge, MA: MIT Press.
- Hasson, U. & Johnson-Laird, P. N. (2002). Belief revision and reasoning. In W.D. Gray & C.D. Schunn (Eds.), *Proceedings of the 24th Annual Conference of the Cognitive Science Society* (p.1009). Mahwah, NJ: Erlbaum.
- Heit, E. (2000). Properties of inductive reasoning. *Psychonomic Bulletin and Review*, 7, 569-592.
- Hewson, C. & Vogel, C. (1994). Psychological evidence for assumptions of path-based inheritance reasoning. *Proceedings of the Sixteenth Annual Conference of the Cognitive Science Society* (pp. 409-414).
- Johnson-Laird, P. N. (1983). *Mental Models: Towards a cognitive science of language, inference and consciousness*. Cambridge UK: Cambridge University Press.
- Johnson-Laird, P. N. (1993). *Human and machine thinking*. Hillsdale, NJ: Erlbaum.
- Johnson-Laird, P. N. & Byrne, R. M. J. (2002). Conditionals: A theory of meaning, inference, and pragmatics. *Psychological Review*, 109 (4), 646-678.
- Johnson-Laird, P. N., Giroto, V., & Legrenzi, P. (2004). Reasoning From Inconsistency to Consistency. *Psychological Review*, 111 (3), 640-661.
- Kim, N. S. & Keil, F. C. (2003). From symptoms to causes: Diversity effects in diagnostic reasoning. *Memory and Cognition*, 31 (1), 155-165.
- Kraus, S., Lehman, D., & Magidor, M. (1990). Nonmonotonic reasoning, preferential models and cumulative logics. *Artificial Intelligence*, 44, 167-207.
- Legrenzi, P. & Johnson-Laird, P. N. (2005). The evaluation of diagnostic explanations for inconsistencies. *Psychologica Belgica*, 45 (1), 19-28.
- Lifschitz, V. (1988). Benchmark problems for foremal nonmonotonic reasoning, version 2.00. In J. Siekmann, *Lecture notes in Artificial Intelligence. Nonmonotonic reasoning* (pp.202-219). Berlin: Springer-Verlag.
- McDermott, D. & Doyle, J. (1980). Non-Monotonic Logic I. *Artificial Intelligence*, 13, 41-72.
- Oaksford, M. & Chater, N. (1998). *Rationality in an uncertain world*. East Sussex: Psychology Press.
- Oaksford, M., Chater, N., & Larkin, J. (2000). Probabilities and polarity biases in conditional inference. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26 (4), 883-899.
- Oshershon, D.N. (1975). Logic and models of logical thinking. In R.J. Falmagne (Ed.), *Reasoning: Representaiton and process in chrildren and adults* (pp. 81-91). Hillsdale, NJ: Erlbaum.
- Pearl, J. (1988). *Probabilistic reasoningin intelligent systems: Networks of plausible inference*. San Mateo, CA: Morgan Kaufman.

- Pelletier, F. J. & Elio, R. (1997). What should default reasoning be, by default? *Computational Intelligence*, 13, 165-187.
- Pelletier, F. J. & Elio, R. (in press). The case for psychologism in default and inheritance reasoning. *Synthese*.
- Pfeifer, N. & Kleiter, G. D. (in press). Coherence and nonmonotonicity in human reasoning. *Synthese*.
- Pfeifer, N. & Kleiter, G. D. (2005). Towards a mental probability logic. *Psychologica Belgica*, 45 (1), 71-99.
- Politzer, G. & Carles, L. (2001). Belief revision and uncertain reasoning. *Thinking and Reasoning*, 7 (3), 217-234.
- Rescher, N. (1964). *Hypothetical reasoning*. Amsterdam: Elsevier, North-Holland.
- Revlín, R. & Calvillo, D. (2005). Counterfactual reasoning: A dual space view. *Manuscript submitted for publication*.
- Revlín, R., Calvillo, D., & Ballard, S. (2005). Counterfactual reasoning: Resolving inconsistency before your eyes. *Psychologica Belgica*, 45 (1), 47-56.
- Revlín, R., Cate, C. L., & Rouss, T. S. (2001). Reasoning counterfactually: Combining and rendering. *Memory and Cognition*, 29 (8), 1196-1208.
- Revlis (1974). Prevarication: Reasoning from false assumptions. *Memory and Cognition*, 2, 87-95.
- Rips, L. J. (1983). Cognitive processes in propositional reasoning. *Psychological Review*, 90, 38-71.
- Rips, L. J. (1994). *The psychology of proof*. Cambridge, MA: Routledge.
- Rumain, B., Connell, J., & Braine, M. D. S. (1983). Conversational comprehension processes are responsible for reasoning fallacies in children as well as adults: IF is not the biconditional. *Developmental Psychology*, 19, 471-481.
- Schroyens, W. & Schaeken, W. (2003). A critique of Oaksford, Chater, and Larkin's (2000) conditional probability model of conditional reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29 (1), 140-149.
- Schurz, G. (2001). What is 'normal'? An evolution-theoretic foundation for normic laws and their relation to statistical normality. *Philosophy of Science*, 68 (4), 476-497.
- Sperber, D. & Wilson, D. (1986). *Relevance: communication and cognition*. Oxford: Blackwell.
- Verhaert, O. (2004). Belief revision and the elimination of inconsistencies. Unpublished Master's thesis. University of Leuven, Belgium.
- Vogel, C. (1996). Human reasoning with negative defaults. In D. Gabbay & H. J. Ohlbach (Eds.), *Lecture notes in artificial intelligence: Practical reasoning* (pp. 606-621). Springer, Berlin.
- Wason, P. C. (1966). Reasoning. In B. M. Foss (Ed.), *New horizons in psychology*, Vol. 1 (pp.135-151). Hamondsworth, UK: Penguin.