HORROR CONTRADICTIONIS

Johan van Benthem, Amsterdam & Stanford, http://staff.science.uva.nl/~johan

Abstract

Contradiction between sharp propositions is a major engine of progress in logic, both in reasoning and in related tasks. And yet, in communicative practice, we often try to avoid conflict, and logic also has several strategies for 'defusing' contradictions. This paper discusses some of these, including changes of arity for predicates, relativization of domains, and retreat to weaker statements about agents' beliefs. We note a few folklore facts about the reach of these methods, and then relate the balance between accepting and defusing contradictions to issues in belief revision and game theory. In doing so, we place 'relativism' in a setting of information dynamics: an optimal take on a contradiction depends on its success in facilitating subsequent communication.

1 Introduction

'Relativism' is a broad topic, and despite Steven Hales' kind invitation to join his team, I am no expert. The philosophical literature that I consulted to upgrade my education (MacFarlane (2009) is an example) has a sophistication in thinking about disagreement, relativism, and contextualism that I simply cannot match. Still, relativism intrigues me as a logician. In practice, many people have found concrete logical inferences universal, and about the only hard currency in cross-cultural communication – whereas others find logical reasoning patterns a culture-laden pawn in the clash of civilizations. And inside the field, I feel a similar tension. Logic mass-produces systems for people's favourite styles of reasoning, a 'relativist' feature in line with Arthur Prior's view of the logician as a lawyer. And yet a major point of logical meta-theory is keeping this diverse field together, finding a common perspective on reasoning styles that look different at first sight. Beneath apparent diversity, one then finds famous translations between rivals like intuitionistic or classical logic, going back to Glivenko, Gödel, and others in the 1930s, making the question of identity and difference of logics a highly non-trivial one. And discoveries of deep identities are continuing. A surprising recent case are convergences between what used to be thought of as major alternatives in reasoning with recursion: modal provability logics and fixed-point logics of computation (van Benthem (2006)).

Diversity and identity for systems of reasoning are grand themes in the philosophy of logic, and they may well have a moral for general issues of relativism. But they are not my concern in this short note. My aim here is much more down-to-earth. Relativism in its colloquial sense seems to pose a threat to a major feature of logic as it is practised: the ability to formulate clear statements that can be true or false, and basic techniques for dealing with the resulting well-defined conflicts. These are the engine of progress. We learn most spectacularly by standing refuted – and also in group settings, the truth emerges from the clash of opinions. But on the other hand, logic is also a source of techniques for 'dissolving' disagreements rather than solving them, as I will show soon with some simple scenarios for this. Eventually, I try to make up my mind: is logic about learning from contradictions and disagreements, avoiding them, or both?

2 Logic and confronting the truth

One might think that logic is about valid proof, and moving from one acknowledged truth to another: with a society of yes-men as the ideal. But this image is historically false, and logic probably originated in a culture of legal, political, and scientific debate. And debates are all about disagreement and refutation, driven by the very same deductive arguments that can also pile truth on truth. Indeed, intuitively, many major notions in logic involve disagreement between parties, which then gets resolved in some way. This becomes especially vivid when we cast things in terms of logical *games*.

Arguing about the truth: evaluation games Resolving clear-cut differences of opinion, while pulling apart different roles, drives games for many logical tasks, as proposed by Lorenzen, Ehrenfeucht, Hintikka, and others (cf. the survey in van Benthem (2008)). A paradigmatic scenario are *first-order evaluation games*, whose bare essentials are as follows. While truth seems just an eternal relation between sentences and the world, the essential underlying process, even in logic, is the dynamics of evaluating sentences. And this is naturally cast in the form of a game. Let two players disagree about a formula φ in some model $\boldsymbol{M}$ with variable assignment s: *Verifier* $\boldsymbol{V}$ claims that φ is true, *Falsifier* $\boldsymbol{F}$ that φ is false. In this setting, natural moves of defense and attack break down complex formulas into components, defining a game $\boldsymbol{game}(\varphi, \boldsymbol{M}, s)$ with

> choices for Verifier when the current formula is a disjunction or an existential quantifier, and choices for Falsifier with conjunctions and universal quantifiers.

> Once an atomic statement Px is reached, one tests if $M, s \models Px$, and
> Verifier wins if this atom is true, while Falsifier wins otherwise.

It is well-known that a first-order statement is true in a model iff Verifier has a winning strategy in the associated evaluation game, while Falsifier has a winning strategy if the formula is false. Thus, indeed truth (or falsity) emerges in the clash of opinions.

Arguing about consistency: proof and model building If no concrete situation is at hand for inspection, as in many conversations, and also scientific reasoning, logic again celebrates crucial episodes of disagreement. When we can derive a contradiction from your and my opinions, we have a clear clash that calls for action. And the contradiction need not even be between human agents: if you observe something that contradicts your current beliefs (Nature's views are inconsistent with yours), you change your beliefs.

But even prior to that, seeing if there is a contradiction in the first place may involve opposite claims. The well-known method of 'semantic tableaux' can be cast as a game between two players. 'Builder' claims the consistency of all assertions made at the start and says that they have a model, while her opponent 'Critic' claims there is no model, and some contradiction can be made manifest. If Builder has a winning strategy, it will generate models of the data, while winning strategies for Critic will be proofs of contradictions from the data. Again, conflicting claims are the trigger, and the basic logical notions of *model* and *proof* emerge in one single setting. Similar roles play in legal procedure between lawyers and prosecutors (as well as judge, jury, or defendant).

More examples could be given, but my point is that well-defined conflict lies at the heart of logic, and it is even useful to seek it in settings where you might not expect it.

But this is only one side of the story. Here is another way of thinking about the above logic games, starting with what may just seem a little cloud on the horizon.

Private interpretation? The above crisp scenarios have unrealistic features when going from formal to natural language. Not all atomic tests in evaluation or discussion games need be clear-cut episodes. Many linguistic expressions contain 'theoretical predicates', such as 'fragile', 'trivial', or 'helpful', whose interpretation leaves room for manoevre. (In Carnap's terms, these are 'dispositional' expressions, in between observation and theory.) Though agents in our disagreement game scenario may share an observational

vocabulary on whose interpretation they agree, they still have great freedom how to interpret these further terms – and an atomic test whether "Mr. *X* is helpful" may well depend on what the 'owner' of this atomic claim at the final stage means by it.

'Open' games This more fluid setting is not a mere nuisance: it may even be of interest to adapt logic games. Tracking rules of defense and attack identifies the player whose claim is at stake at any stage. (Positive occurrences of atomic sub-formulas in the initial formula, lying under an even number of negations, would be owned by the Verifier, and negative ones by Falsifier.) In particular, in a free-standing atomic formula *p*, Verifier's claim is at stake, and it seems reasonable to let *her* interpretation of *p* count in the test. By contrast, in a free-standing negation ¬*p*, a role switch has taken place, and Falsifier now owns the claim, viz. that *p* is false: so he has a right to determine the interpretation. In addition to such 'shaky atoms', players may still agree on the interpretation of part of the vocabulary, of course. Such extended evaluation games are still related to the earlier games, now for modified formulas, substituting 'True' for shaky atoms owned by Verifier (assuming that she will choose interpretations making her win), and 'False' for those owned by Falsifier. There are more intrinsic new features, too. But my point here is just that 'simple' logical scenarios of disagreement may not be so simple – even though we can adapt our game modeling, and logic does not go out of the door.

'Open' debates? Likewise, one can cast debates so as to allow for a special role of the theoretical vocabulary. Here is an example of additional freedom for agents, just to show a tactic different from the above one. Distinguishing shared vocabulary $\boldsymbol{P}$ from freely interpretable vocabulary $\boldsymbol{Q}$, a clash of opinions between prima facie contradictory assertions $\varphi(\boldsymbol{P}, \boldsymbol{Q})$ and $\neg\varphi(\boldsymbol{P}, \boldsymbol{Q})$ now becomes a consistency problem for the formula

$$\exists \boldsymbol{Q}\, \varphi(\boldsymbol{P}, \boldsymbol{Q}) \wedge \exists \boldsymbol{Q'}\, \neg\varphi(\boldsymbol{P}, \boldsymbol{Q'}).$$

But here we find a surprise: logical theorems may bear on conversational issues. Theory-laden clashes in this format add no disagreements beyond the base level:

> If $\exists \boldsymbol{Q}\, \varphi(\boldsymbol{P}, \boldsymbol{Q}) \wedge \exists \boldsymbol{Q'}\, \neg\varphi(\boldsymbol{P}, \boldsymbol{Q'})$ is inconsistent, $\exists \boldsymbol{Q}\, \varphi(\boldsymbol{P}, \boldsymbol{Q}) \rightarrow \forall \boldsymbol{Q'}\, \varphi(\boldsymbol{P}, \boldsymbol{Q'})$ is valid. But then, by the *Interpolation Theorem* for first-order logic, there is a formula $\alpha(\boldsymbol{P})$ in the $\boldsymbol{P}$-vocabulary only, with $\models \varphi(\boldsymbol{P}, \boldsymbol{Q}) \rightarrow \alpha(\boldsymbol{P}) \rightarrow \varphi(\boldsymbol{P}, \boldsymbol{Q'})$. But then $\exists \boldsymbol{Q}\, \varphi(\boldsymbol{P}, \boldsymbol{Q})$ implies $\alpha(\boldsymbol{P})$, while $\exists \boldsymbol{Q'}\, \neg\varphi(\boldsymbol{P}, \boldsymbol{Q'})$ implies $\neg\alpha(\boldsymbol{P})$.

Thus, $\alpha(\boldsymbol{P})$ versus $\neg\alpha(\boldsymbol{P})$ is still a base-level disagreement between the two parties. Still, clearly, $\varphi(\boldsymbol{P}, \boldsymbol{Q}) \wedge \neg\varphi(\boldsymbol{P}, \boldsymbol{Q'})$ may well have become consistent – in which case the prima facie disagreement has gone away. This 'predicate splitting' will return below.

3 Avoiding contradiction in discourse

Is 'diluting confrontation' an epicycle to standard logical patterns? Or is more at stake? Moving from formal languages to ordinary discourse, real logical clashes seem rare. Or at least, there is a whole array of mechanisms for fast pain relief. If you contradict me, we may quickly decide to compromise, each retreating a little. (My American dentist claimed that my mother tongue Dutch is not a real language. Eventually, we agreed on the compromise that she offered: "OK, Johan, you speak it, but nobody *writes* it".) The compromise can also be softer, in terms of the context dependence studied by linguists and philosophers. I find the Dutch socialist elite rich, given my small reference group of affluent people, but you may find them poor. This is inevitable: context dependence of meanings is so prevalent in natural language, that sentences seem to say more about how a speaker uses her language than about the world. Add evaluative language, vagueness, and other key features of ordinary language and communication – and it is a miracle that we manage to disagree at all. And the repertoire of clash avoidance is still richer. We also have the 'meta-ability' to look at statements devoid of content, merely being on the table, but as yet without any active attitude on our part. As a department chair in Amsterdam, I often had to exercise this ability when dealing with distinguished professors. (Since our editor does not allow any footnotes, I cannot provide juicier details here.) Language is largely a cooperative game, and just as early science had its 'horror vacui', much of communication seems driven by 'horror contradictionis'. Even public academic debates are often boring, the compromising starts right on the podium.

In this light, maybe we have found a somewhat unusual but important task for logical analysis of natural language: assisting people in achieving *genuine disagreement*? There are precedents for this view. In the early 1950s, Stalin attacked relativism in linguistics and logic by observing that, if language were class-relative, Marxism would be undermined. Class struggle would be just talking at cross-purposes. This line helped east-bloc logicians survive in the difficult 1950s. But can logic really deliver?

4 Relativism inside logic: ways of avoiding contradictions

Things are not so clear-cut. Pressure to remove contradictions in a way that gives every person (or every consideration) something is also felt inside logic, and I will discuss this theme in the rest of this paper as my modest take on 'relativist' tendencies. In fact, there are several ways in which this can be done, so logic even helps the relativist.

To see this, ask yourself: what *does* logic do with contradictions? To be sure, there is an ideal Popperian 'learning response'. Belief revision theories (Gärdenfors & Rott (1995)) note a contradiction, accept it as stated, and then revise in Ramsey's terms: "give up enough old beliefs to accommodate the favoured new proposition contradicting them".

But this response is not the only one, and perhaps not even the prevalent one.

Para-consistency Perhaps the most striking trend today that refuses to be budged by a contradiction is the use of a para-consistent logic that tolerates inconsistencies. This is often motivated by an appeal to real behaviour: our beliefs are full of inconsistencies, but as long as we compartmentalize things, we can still draw a lot of useful inferences, and there is no need to throw away an inconsistent theory. While I agree with the latter observation, I do not think it follows that we need a logic that tolerates *manifest* contradictions, once they have come to our attention. What will ever jolt our creative imagination, if not a contradiction? But I admit life is easy on a para-consistent atoll: one can be informed of the most blatant contradictions, tune out, and happily turn to something else – even without a minimal relativist effort to patch up things somewhat.

Giving up, and holding on But then, I admit that I have never felt great resonance to the para-consistent stance. To me, contradiction is a major engine of progress in logic. But just how? On the simplest account, that we all know, the response is easy. We have been too greedy. The sum total of all our beliefs is untenable, and we must now retreat, saving chunks of the earlier theory, hopefully large useful ones. This pattern often makes a lot of sense when adjusting our beliefs to reality, and indeed, it underlies the state change driving modern belief revision theories (though more on that below).

But responses can be much more sophisticated. Think of the medieval recommendation *"In case of a contradiction, make a distinction",* that 'repairs' contradictions instead of accepting them at face value. I will discuss a few ways in which this works, making strategies for relativism more concrete. Everything I have to say is quite elementary.

My cues come from the neglected gem Weinberger (1965), an original study of repair mechanisms, in logical practice, but also in the history of science. I will discuss two.

Arity Raising One major line, reflecting modern contextualism, is the 'Arity Raising Strategy': "give some predicate a new argument that can vary in the given assertions". What is helpful to you need not be helpful to me, what is true in your pragmatic context need not be true in mine, and so on. Examples also occur in science, as Weinberger points out: to the Eleatics, things both moved and did not move, but physics added the parameter of a 'frame of reference'. Technically, this strategy takes one or more k-ary predicates $Px_1 \ldots x_k$ in the language, and raises the arity by one: $P\mathbf{y}x_1 \ldots x_k$ with an extra argument position $\mathbf{y}$, for either old or new objects. A formula φ in which P occurs then gets *arity-raised variants* φ' where we can set the values of the free variable y differently at different occurrences. This ploy suffices for removing any contradiction in the most common logical systems. What follows are probably folklore results – but they still see worth stating to get some basics out in the open:

Fact In propositional logic, any formula φ has a consistent arity-raised variant, where only two objects are needed to distinguish predicates.

Proof By standard logic, each occurrence of a proposition letter p in a formula is either 'positive' or 'negative', following a standard syntactic recursive definition (e.g., occurrences in components of a conjunction retain their 'polarity', negations 'switch polarity', etc.) Now replace all negative occurrences of p in φ by some new atom p', making all occurrences of p positive, and al occurrence of p' negative. Then, the atomic valuation making all original atoms p true and all variant p' false makes the whole φ true. This follows by a simple induction on formulas. (Likewise, making all p false and all p' true makes φ false.) We can think here of the p as referring to one situation, with p' referring to another. There are also other ways of getting this result that may look less ad-hoc. For instance, one can 'open up' any closed branch in a semantic tableau for φ and remove overlaps between proposition letters on its left and right by priming, and then extend the marking appropriately to all complex formulas on the branch. ■

Clearly, this is extremely simple non-sophisticated logic. But that is what I told you!

The tableau method with priming conflicting predicates also applies to predicate-logic, but transferring markings to complex formulas is less easy. Still, say, the contradiction $\exists x \forall y(Rxy \leftrightarrow \neg Ryy)$ underlying the Russell Paradox becomes consistent in the variant

$$\exists x \forall y(Rxy \leftrightarrow \neg R'yy).$$

More generally, the earlier drastic method of marking predicates still works:

Fact Each predicate-logical formula φ becomes consistent when all occurrences of predicate letters P are made disjoint, in variants $P, P', P'', \ldots$

Proof Without loss of generality, move negations in φ inside until they stand in front of atoms. This does not multiply occurrences of atoms, and the result is an equivalent formula built from literals (atoms and negated atoms) using only conjunction, disjunction, universal and existential quantification. Now consider a one-object model, where both quantifiers amount to just the identity operation. Make all predicates with positive occurrences of their single atom true, and all others false. This makes all atoms and negated atoms in the formula true, and the quantifiers, conjunction, and disjunction in φ do not change truth, so the whole compound formula is true as well. ■

Of course, this is overly drastic in real settings, and it is an interesting technical issue which *minimal* forms of 'surgery' will make a given formula consistent. But anyway, the point of these results is not that they are realistic methods in practice. They just show that contradictions can always be removed. And once you have seen that in principle, more intelligent versions can be found. And as a further point, we see that logical methods like semantic tableaus are a two-edged sword. They can be used to detect inconsistencies, but for that very reason, they can also be used to avoid them!

Domain Restriction While Arity Raising is a context-related tactic, Weinberger also discusses a second method, reflecting another ubiquitous phenomenon in natural language, viz. the implicit *domain dependence* of quantifiers. Even when not explicitly marked, quantified noun phrases come with restrictions on their range. When I say that "No student passed the exam" I always mean students from some relevant domain. Indeed, there is a whole semantic literature on how such implicit quantifier domains can shift inside and across sentences when we use natural language (Westerståhl (1984)).

In particular, for any first-order formula we can define *domain variants* by restricting one or more quantifier occurrences to some syntactic predicate, old or new. This, too,

corresponds to real conversational and even scientific strategies. For instance, consider Cantor's original set theory, whose unrestricted Comprehension Principle turned out inconsistent. While Zermelo-Fraenkel set theory uses a 'weakening strategy' limiting Comprehension, Von-Neuman-Bernays-Gödel set theory removes the contradiction by distinguishing quantifiers ranging over two domains: 'sets' and 'classes'. For a concrete illustration, take again the above Russell Formula $\exists x \forall y(Rxy \leftrightarrow \neg Ryy)$. Without using Arity Raising, another consistent variant is $\exists x\,(Ax \wedge \forall y\,(By \rightarrow (Rxy \leftrightarrow \neg Ryy))$.

Domain Restriction is not as general as Arity Raising. It cannot defuse hard-core propositional contradictions, witness the first-order formula $\exists x(Px \wedge \neg Px)$ that remains a contradiction if we just relativize the existential quantifier. Here is what it does achieve. Without loss of essential structure, rewrite a formula φ as before to a compound of quantifiers, conjunction, and disjunction over (negated) atoms. Looking outside in, this is a compound of (negated) atoms plus sub-formulas prefixed by universal quantifiers, using only conjunction, disjunction and existential quantifiers. Now replace each universal sub-formula in φ by *True* to obtain a formula φ^*. (The idea in what follows is that universal formulas can always be made true trivially by restricting their initial quantifier to the empty domain.) Then, turn the formula φ^* into a purely propositional φ^{**} by substituting distinct individual constants for each existential quantifier.

Fact A first-order formula φ has a consistent domain variant iff φ^{**} is consistent.

Proof If some domain variant of φ has a model, then by a simple induction on the above construction rules (conjunction, disjunction, existential quantification), the variant φ^* is true in that same model. But then, taking witnesses for the existential quantifiers, φ^{**} is also consistent. Vice versa, if φ^{**} holds under some propositional valuation, we can form a model for φ^* with just the objects mentioned in the atoms. The earlier trick then gives a model: replace universal sub-formulas $\forall x\alpha$ by the always true $\forall x(False(x) \rightarrow \alpha)$, while we do not even have to relativize the existential quantifiers. ■

As an illustration where this does no good, $\exists x(Px \wedge \neg Px)^{**} = Pd \;\&\; \neg Pd$, and the latter shape explains why we do not get consistency here. Of course, even in cases where we do get consistency, this is not a realistic method, and the restriction to empty domains is a trick that is even less informative than the formal Russell variant mentioned earlier. Putting more constraints on admissible transformations of the original formula would

raise interesting new issues. My point with the preceding trivial fact is just that Domain Restriction, too, has wide scope in avoiding contradictions – and logic can tell us how.

Conclusion Logic has mechanisms for dealing with contradictions, and in principle, they can remove all of them. Here are two extreme methods. The first is to accept the contradiction as stated, and change our present theory. This is the driving force of most logics of belief revision and learning generally. The other extreme is to defuse the contradiction by some intelligent variant of the methods that we have outlined. This, too, will have repercussions for the ways in which we phrase our current theories, and authors like Hans Rott have stressed the open problem how to merge the two extremes: acceptance and language change into one account of how we learn (cf. Parikh (2009)). So, we end up somewhere in the middle: logic is about sophisticated truth clashes.

5 Avoiding contradiction in the setting of agency

Our conclusion in the preceding section refers to beliefs. And this suggests an extension of our discussion going beyond Weinberger (1965). It takes two to disagree, and hence *agents* enter the picture. Then we must move to logics of knowledge or belief, including recent dynamic versions. Here is my favourite setting for disagreement scenarios:

Belief juxtaposition In logics of agency, a third mechanism of defusing contradictions arises naturally. Its point of entry is this. When people make assertions φ, there is a long-standing issue of just what they are saying. Do they know the stated proposition φ, do they only believe it, or in between, do they perhaps believe that they know? The latter agent-relative information is not just a ploy: serious communication scenarios in logic, computer science, and game theory are driven by epistemic information, not about base facts, but what other agents know or believe (Geanakoplos & Polemarchakis (1982), Fagin, Halpern, Moses & Vardi (1995), van Benthem (to appear)).

We cannot really relativize by saying that one agent knows that φ and the other that $\neg\varphi$, since this would leave the contradiction intact, now in the veridicality of knowledge. But of course, the *retreat to beliefs* is a powerful way of avoiding contradictions.

Fact If both φ and $\neg\varphi$ are consistent propositional formulas, then so is $B_1\varphi \wedge B_2\neg\varphi$.

Proof Just take the space of all propositional valuations, and add any plausibility order for 1 that puts the φ-worlds on top, and for 2 that puts the $\neg\varphi$-worlds on top. ■

Note that, in a sense, this move still exemplifies both earlier strategies, as we relativized φ, $\neg\varphi$ to different sets of worlds, so we add a parameter, and also restrict domains. We will focus on the agency aspects per se from now on, leaving this technical aspect aside.

Also, note that this construction does not work with 'inner inconsistencies'. If one of φ, $\neg\varphi$ is a contradiction (say, the agents dispute the truth of some mathematical law), then we need more finely-structured belief models – or, one of our earlier tactics for first defusing the internal contradiction in what was communicated by one of the agents.

Taking agency seriously The preceding observation was just a simple truism. But the agent setting is suggestive. Our focus in the preceding section has been on removing internal contradictions inside one theory, perhaps owned by a single agent. But such contradictions may be due to over-ambition or faulty reasoning, and there need not be any good reason for syntactically defusing them into something consistent. By contrast, relativism comes up most naturally when several agents meet, and have to adjust conflicting beliefs leaving each their dignity. (Of course, one can then reformulate this conflict as a contradiction in the single merged theory, but let's go slowly.)

What natural scenarios of interaction arise in this setting? Game theorists have had the most to offer here, in the wake of the seminal paper 'Agreeing to Disagree' (Aumann 1976). For a start, what is the *point* of belief juxtaposition? One aim might be to arrive at a situation where it is common knowledge in the group *{1, 2}* that the agents have conflicting beliefs. But then, 'Aumann's Theorem' says that this can only happen when the agents started already with different priors: or qualitatively, different over-all plausibility relations on worlds. That is precisely what we created above: the agents disagreed on their plausibility ranking of worlds. But other things can happen, too. The game-theoretic style of analysis includes higher beliefs about beliefs of other agents, and passing information about conflicts of opinion can have quite unexpected effects. We will now make a few points about these in a more logical perspective:

'The next stage': relativism and information dynamics Achieving consistency is only the beginning of a story. We do this out of a further *interest* (otherwise, why bother?): say, to facilitate a subsequent conversation, or even further research. Language and communication are all about successive assertions that modify the context, domains of discourse, and perhaps even arities of predicates describing events at various levels of detail. And that would also be our general take on the importance of understanding

relativism. Dealing with disagreement is not of much interest per se, as a 'static' issue between propositions. But it is important as an issue in the *dynamics of information flow and communication* between agents and communities. But then, the real test of a take on relativism is whether it is successful in initiating further discourse dynamics.

This theme is found in 'dynamic-epistemic logics' of hard and soft information, that study how agents gather and exchange information, and update knowledge and beliefs accordingly (Baltag & Moss (2004), van Ditmarsch, van der Hoek & Kooi (2007), Baltag & Smets (2006), van Benthem (to appear)). Technically, this dynamics changes epistemic-doxastic agent models by eliminating worlds and/or modifying epistemic accessibility and doxastic plausibility relations. In such a framework, the key issue is identifying *which dynamic acts* are crucial to an agent scenario. In fact, the above belief juxtaposition already contained two of these, reflecting intriguing dynamic scenarios that have attracted attention in recent years. Both have to do with *conflicts in beliefs*.

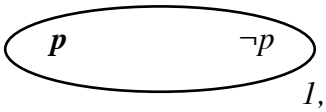
Belief merge When agents meet, with their knowledge and beliefs encoded in private epistemic-doxastic models, then, to get fruitful interaction going, they must *merge* into a shared group model. There is no consensus in the literature on an optimal mechanism for such a merge, and I just gave a simple propositional case. The general desideratum here is a shared model that leaves people their beliefs as far as possible, and that still works in the appropriate language with (iterated) belief operators for both agents.

Belief update But merging is just a start. Once a shared model is in place, further information exchange can start. In that second phase, initial beliefs of agents may change, depending on how reliable one takes others to be. Dynamic-epistemic logics have many mechanisms for this (cf. van Benthem (to appear)), and more are emerging. Baltag & Smets (2009) propose a framework for studying the long-term learning effects of announcing differences in beliefs. Closer to the above-mentioned game theoretic line, Dégrémont & Roy (2009) show how agents can 'synchronize' their beliefs optimally by repeatedly announcing their differences of opinion, as long as they still exist – since the very statement of such differences may be informative. (The mechanisms can be either 'hard' elimination of worlds contradicting the stated information, or 'soft' changes in agents' plausibility relations.) In particular, quite generally, repeated announcement of a conflict in beliefs between two agents will lead to one of two stable situations: (a) that difference becomes common knowledge, and the priors of the agents are different, or

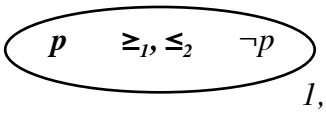
(b) that difference finally disappears, and the agents have come to agree. (The forthcoming dissertation Dégrémont (to appear) even has delightful scenarios where the disagreement can 'flip-flop' in successive rounds, inverting who believes what.)

Concrete scenarios for communicating disagreement While these general results lead us too far, this setting does provide concrete pictures worth pondering. Much depends on the initial situation agents are in: and we have acts of creating plausibility, or once an epistemic-doxastic model is in place, further acts of communication that make information flow. Thus, the dynamics of communicating disagreement comes in many different flavours. I conclude with a few samples just to show how things get more interesting than in my earlier bare discussions in propositional or predicate logic. We start with a simple case where both agents have the same *hard information*:

Scenario 1: Neither agent knows if p, and their plausibility relations are unknown:



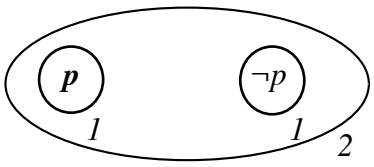
Announcing that $B_1p \wedge B_2\neg p$ may be seen as a soft *plausibility-introducing upgrade* to a situation with different priors for the two agents:



The result is that the disagreement $B_1p \wedge B_2\neg p$ has become common knowledge.

But things can be more complicated when agents have different hard information, and in that case, communicating disagreement may actually resolve the conflict in beliefs:

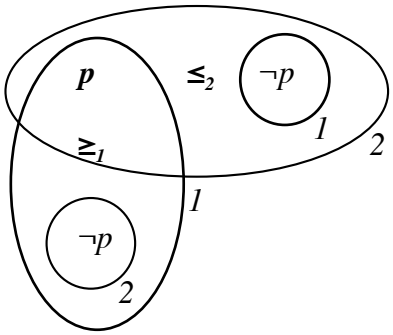
Scenario 2: Agent 1 knows that p is true in the actual world to the left, but *2* does not – and the following picture is common knowledge:



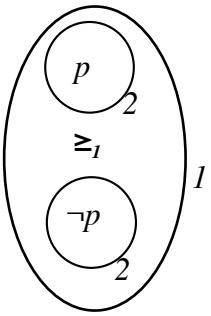
Now $B_1p \wedge B_2\neg p$ cannot become common knowledge by a soft update introducing a plausibility relation that respects the given epistemic accessibility relation. The reason is that *1* will never believe that p in the world to the right. But this difficulty arises

because *1* has hard information that *2* still lacks. Thus, the agents need to communicate first to get into a better epistemic-doxastic balance. Indeed, *communicating the very disagreement* will help both: $B_1 p \wedge B_2 \neg p$ can be true in the actual world only, because, no matter how we take the plausibility relation, *1* can only believe p in that world. Thus, the model reduces to the single actual world to the left, with both equally informed. The result is of course mutual agreement, and indeed this is common knowledge.

Scenario 3: Here is a more complex initial setting, with further initial beliefs present. In the actual world on the top left, *2* believes that $\neg p$, but she does not know what *1* believes about p (to the right, he believes that $\neg p$, while to the left, he believes that p but does not know p) and hence she does not know *whether they disagree* about p:



Stating the true disagreement (true in the actual world) $B_1 p \wedge B_2 \neg p$ reduces this to



Now the beliefs of *1, 2* coincide, since both are directed to the actual world where p holds. Hence, communicating disagreement in belief is not going to help them further, but *communicating agreement* on p would help, as this rules out the bottom world.

This shows that we are not studying one process of belief merge and communication: that would be an elusive quest. We must understand the general phenomenon, including the fact, known from the philosophical literature, that *1 plus 1 = 3* in this setting. When two agents meet, *three actors* result: the separate agents, and the group of them both. And we also see our earlier point demonstrated: disagreement and agreement become

rich phenomena in such a group setting, far beyond single-minded takes on relativism. What I am *not* saying here is that dynamic epistemic logic solves all dynamic aspects of relativism in beliefs, but rather, how it enables us to see more of its full wealth.

Update and language change: re-interpretation once more Indeed, our original problems have not disappeared. All these scenarios of communicating disagreement or agreement get more complicated when agents become aware of their *language* – either because they must merge private languages, or because they worry whether they mean the same by what seemed shared expressions. The earlier mechanisms of linguistic *re-interpretation* then enter the assessment of disagreement after all. Here we hit a boundary, and an open problem that was mentioned earlier returns. Current theories of belief revision or dynamic-epistemic logics have no systematic account of the mix of *plausibility adjustment* and *language change* in realistic knowledge update and belief revision. (For some initial thoughts, cf. Rott (2008), Parikh (2009), and van Benthem (to appear).) This topic is beyond my scope here. But I do claim that logical discussions of relativism in terms of disagreement and contradiction removal strategies remain sterile, if one does not study the plurality of *merge operations*, and after that, take the next step of tackling the logical dynamics of the ensuing *communication*.

6 Conclusion

I started out by saying that logic thrives with clear-cut opinions and disagreements. But I then pointed out that the very tools developed for those purposes of clarity and dissent can also be used to defuse disagreements, and provide logical methods for what we often do in daily practice: walk around conflict. I find it hard to take a crisp stance. Total accommodation of differences in claims stifles progress, and makes a mockery of learning. But taking every contradiction at face value is the sign of an uncultivated mind. Where is the middle ground? Formal logic serves both extremes, as well as mixtures: but finding the right balance remains a matter of good 'rational sense'.

I have also tried to shift the discussion a bit. Relativism as a general concern sounds far-fetched at times, but relativism as a label for dealing with conflicts in communication is highly realistic. I have proposed a broader stance on the dynamics of communication, including processes of merging beliefs and further information exchange afterwards. While this makes the theme richer, including links to game theory, a similar tension

returns. On the one hand, dynamic epistemic logics and formal learning theories follow the line that observations contradicting our beliefs are the engine of progress. But on the other, there are respectable logical mechanisms that can soften the blow, and make new evidence consistent with old beliefs. Finding the right mixture seems important to understanding real conversation and argument, but how to do it in a natural manner?

Finally, to return to an issue raised briefly in the Introduction, language change tactics naturally lead to the theme of *translation*. If you think that my true *p* is your false *p'*, we can translate your discourse into mine. Underneath surface differences in opinion, there can be invariances modulo translation. I leave this theme to another occasion, since this paper has already become quite long for someone who promised he had little to say.

References

Aumann, R. (1976). Agreeing to disagree. *The Annals of Statistics* 4:6, 1236–1239.

Baltag, A. and Moss, L. (2004). Logic for epistemic programs. *Synthese* 139:2, 165—224.

Baltag, A. and Smets, S. (2006). Conditional doxastic models: a qualitative approach to dynamic belief revision. In G. Mints & R. de Queiroz Eds). *Proceedings of WOLLIC 2006*, Springer Electronic Notes in Theoretical Computer Science 165.

Baltag, A. and & Smets, S. (2009. Learning by questions and answers: from belief-revision cycles to doxastic fixed points. In *Proceedings WoLLIC*, Tokyo 2009.

Benthem, J. van (2006). Modal frame correspondences and fixed-points. *Studia Logica* 83:1, 133 – 155.

Benthem, J. van (2007). Logic games, from tools to models of interaction. In A. Gupta, R. Parikh and J. van Benthem (Eds). *Logic at the Crossroads*, Mumbai: Allied Publishers, 283–317.

Benthem, J. van (to appear). *Logical dynamics of information and interaction*. Cambridge: Cambridge University Press.

Dégrémont, C. (to appear). *Wisdom comes with age, the dynamics of belief over time*, Dissertation, Marie Curie Centre 'Gloriclass', Institute for Logic, Language and Computation, University of Amsterdam.

Dégrémont, C. and Roy, O. (2009). Agreement theorems in dynamic-epistemic logic. Stanford: *Proceedings TARK 2009*.

Ditmarsch, H. van, Hoek, W. van der, and Kooi, B. (2007). *Dynamic epistemic logic*. Dordrecht: Springer.

Fagin, R., Halpern, J. Moses, Y. and Vardi, M. (1995). *Reasoning about knowledge*. Cambridge (Mass.): The MIT Press.

Gärdenfors P. and Rott, H. (1995). Belief revision. In D. M. Gabbay, C. J. Hogger and J. A. Robinson (Eds). *Handbook of logic in artificial intelligence and logic programming* 4. Oxford: Oxford University Press.

Geanakoplos, J. and Polemarchakis, H. (1982). We can't disagree forever. Cowles Foundation Discussion Papers 639, Cowles Foundation, Yale University, July 1982.

MacFarlane, J. (2009). Varieties of disagreement. Department of Philosophy, University of California, Berkeley.

Parikh, R. (2009). Beth definability, interpolation and language splitting. New York: CUNY. To appear in J. van Benthem, Th. Kuipers and H. Visser (Eds). Proceedings Beth Centenary Symposium 2008, *Synthese*.

Rott, H. (2008). Information structures in belief revision. In P. Adriaans and J. van Benthem (Eds). *Handbook of the philosophy of information*. Amsterdam: Elsevier, 457 – 482.

Weinberger, O. (1965). *Der relativisierungsgrundsatz und der reduktionsgrundsatz – zwei prinzipien des dialektischen denkens*. Prague: Nakladatelství Ceskoslovenské Akademie Ved.

Westerståhl, D. (1984). Determiners and context sets. In J. van Benthem and A. ter Meulen, (Eds). *Generalized quantifiers in natural language*. Dordrecht: Foris, 45–71.

JOHAN VAN BENTHEM

University Professor of Logic, University of Amsterdam. Henry Waldgrave Stuart Professor of Philosophy, Stanford University. Weilun Professor of Humanities, Tsinghua University, Beijing. Current main interests: logical dynamics of information flow and communication, logic and games, logic and epistemology.