Belief Merging Based on Adaptive Interaction

Giuseppe Primiero

1 Introduction

Adaptive Logics ([Bat01], [Bat07]) are designed to handle forms of internal dynamics, by a better comprehension of the premise set at each stage of a derivation, as well as of external dynamics by the addition of new premises. They provide a new approach in the standard research programme of Dynamic Doxastic Logic ([AGM85], [vB96], [vD05], [LS07]), by enabling one to formalize information merging in a multi-agent setting. The logics $\mathbf{ADM}^r$ and $\mathbf{ADM}^m$ (*Adaptive Doxastic Merging* with *Reliabilty* and *Minimal Abnormality* Strategies – simply referred to as **ADM** when properties common to both logics are considered) are able to deal with the case of some belief content holding in a certain belief state ($b_1\phi$) and rejected in some other belief state ($\neg b_2\phi$). A belief is merged if it is consistently deduced in one agent's belief state and is coherent with the contents of any other agent's belief state. The interaction for merging is determined by the different strategies, providing a machinery to localize the contents which cannot be granted to be acceptable for all the interacting agents. The present model might be seen as a multi-agent formulation of the notion of "being informed" presented in [All06]. The informal meaning of the two strategies can be presented as follows: *reliability* accounts for complete consensus between the agents, avoiding global dissatisfaction as much as possible; *minimal abnormality* accounts for partial disagreement, minimizing individual dissatisfaction as much as possible. The strategies are therefore proper counterparts to the two standard merging operators of *arbitration* and *majority* ([LS98], [LM99], [KPP02]). Reliability is an extremely cautious interaction, which also means that information may be deleted which might have been safely derivable by a single agent. Minimal Abnormality formulates restrictions on the contents which cannot be accepted by the group in its entirety, but at least by some of the participants: in other words, it resolves the inconsistencies by allowing groups of agents to form internal alliances.

2 The Lower Limit Logic DM and the Upper Limit Logic M

The language $\mathcal{L}^B$ of the *Lower Limit Logic* **DM** of **ADM** is obtained by extending the standard language $\mathcal{L}$ of classical logic (**CL**) with a set of doxastic belief operators of the form b_i (for every $i \in \mathcal{I} = \{0, 1, \ldots, n\}$). Intuitively, $b_i\phi$ (for $i > 0$) will express that agent i believes ϕ; the operator b_0 will be used for the beliefs that belong to the merged state. The set of well-formed formulas $\mathcal{B}$ is restricted to well-formed formulas of $\mathcal{L}$ and modal well-formed formulas of first degree. The set of primitive well-formed formulas is referred to as $\mathcal{B}^P$, the set of atoms (propositional letters and their negations) as $\mathcal{B}^A$ and the set of disjunctions of one or more atoms as $\mathcal{B}^\vee$. The later needed notion of "primitive beliefs" is a disjunction of atoms. Capital letters from the beginning of the alphabet will be used as metavariables for members of $\mathcal{B}$; letters from the greek alphabet $\phi, \psi, \ldots$ are used as metavariables for non-modal formulas. A **DM**-model is a quadruple: $M : \langle \mathcal{W}, w_o, \mathcal{R}, v \rangle$, in which $\mathcal{W}$ is a set of worlds and $w_o \in \mathcal{W}$; $\mathcal{R}$ is a set of accessibility relations $R_i : \{w_0\} \to \mathcal{W}$ ($i \in I$), and $v : \mathcal{B}^P \times \mathcal{W} \to \{0, 1\}$ an assignment function. The valuation v_M defined by the model M is characterized in a standard way by defining the connectives $(\neg, \wedge, \vee, \supset)$. A doxastic formula in such a model is defined as follows: $v_M(b_i\phi, w) = 1$ iff $v_M(\phi, w') = 1$ for all w' such that $R_i ww'$. A model M verifies A iff $v_M(A, w_0) = 1$, $\Gamma \models_{\mathbf{DM}} A$ iff all **DM**-models of Γ verify A, and $\models_{\mathbf{DM}} A$ iff all **DM**-models of Γ verify A. **DM** is thus a **KD4**-style logic restricted to first degree (indexed) operators. **DM** is axiomatized by extending a propositional fragment of **CL** with the standard axioms of Monotony, Consistency and Necessitation. The *Upper Limit Logic* for **ADM** is called **M**, obtained from **DM** by stipulating that, for all $i, j \in \mathcal{I}$ and for all $w, w' \in \mathcal{W}$, $R_i ww'$ iff $R_j ww'$. This entails that for every ϕ and for every $i, j \in \mathcal{I}$, $b_i\phi \supset b_j\phi$. According to this restriction, if one agent holds a belief, any other agent interacting with him will accept that belief. This property expresses what can be called a *merging condition*: $b_i\phi \supset b_j\phi$ for all $i, j \in \mathcal{I}$. Belief contents for which this condition does not hold will represent conflicts among the agents involved in the interaction. The **ADM**-consequence set of a set of premises Γ will extend the **DM**-consequence set of Γ by presupposing that as many of these conflicts as possible are false.

3 Interaction by Abnormalities

The role of the strategies in **ADM** is to localize the conflicting beliefs. The formalization of the non-mergeable beliefs is given by the definition of the so-called *abnormalities*. Their logical form in **ADM** expresses that a certain content is believed by agent i ($b_i\phi$), but disbelieved by agent j ($\neg b_j\phi$). This turns out to be the negation of the previously mentioned *merging condition*: $b_i\phi \wedge \neg b_j\phi$. Abnormalities in **DM** therefore express a *merging failure*. The set of abnormalities Ω contains all the formulas expressing that an agent's belief is not a mergeable belief, restricted to primitive beliefs: $b_i\phi \wedge \neg b_0\phi$ ($i \in \mathcal{I}$ and $\phi \in \mathcal{B}^{\vee}$). Whereas a formula of the form $b_i\phi \wedge \neg b_0\phi$ is not **DM**-derivable from any set of premises Γ, disjunctions of such formulas may be derivable. A disjunction of abnormalities will be called a *Dab*-formula. If Δ is a singleton, $Dab(\Delta)$ is simply an abnormality, i.e. a member of Ω. If Δ is empty, $Dab(\Delta)$ is empty as well. The *Dab*-formulas derivable by **DM** from Γ will be called *Dab*-consequences of Γ ($\Gamma \models_{\mathbf{DM}} Dab(\Delta)$). $Dab(\Delta)$ is a *minimal Dab-consequence* of Γ if there is no $\Delta' \subset \Delta$ such that $\Gamma \models_{\mathbf{DM}} Dab(\Delta')$. From each set of premises Γ a number of minimal *Dab*-consequences may be derivable, each containing the assertion of one or more beliefs which cannot be merged. The core idea of the adaptive logic is to avoid *Dab*-consequences "as much as possible".

4 The Semantics of ADM^r and ADM^m

For some premise set Γ, some $Dab(\Delta)$ with Δ not a singleton, will be derivable by **DM**. Because it is not established *which* member of one such Δ is in fact abnormal with respect to Γ, the role of the strategy is to establish what it means to interpret Γ as normally as possible. In other words, the **ADM**-consequences are all those formulas which can be derived from Γ by **DM**, by supposing that the members of Ω defined in **DM** are false, insofar this is possible. To establish when this is not possible is the role of the strategy. If Γ requires no abnormality to be true, it will have models in **M**; a model of **M** is a model of **DM** which verifies no abnormalities. Each strategy provides different selections on models and different consequences. Strategies might be interpreted as the ways the agents interact in order to overcome conflicting beliefs.

Reliability Strategy. The formulation of this strategy is based on the definition of unreliable formulas of a model. Given all the minimal *Dab*-consequences of a premise set Γ ($Dab(\Delta_1), Dab(\Delta_2), \ldots$) the set of unreliable formulas with respect to Γ is given by their union ($U(\Gamma) = \Delta_1 \cup \Delta_2 \cup \ldots$). The reliability of a **DM**-model of Γ is given by the relation between the unreliable formulas and the so-called abnormal part of a model. Where M is a **DM**-model, the abnormal part of M refers to the set of abnormalities verified in M ($Ab(M) = \{A \in \Omega \mid M \models A\}$). The semantic selection performed by the Reliability strategy establishes that all abnormalities verified by a model are the unreliable formulas of the given premise set. Hence, according to this strategy a **DM**-model M of Γ is a reliable model of Γ iff $Ab(M) \subseteq U(\Gamma)$. The consequence relation for **ADM**^{r} is defined with respect to the reliable models: for any $A \in \mathcal{B}$, $\Gamma \models_{\mathbf{ADM}^r} A$ iff A is verified by all reliable models of Γ. According to Reliability, the different agents are maximally cautious: as soon as they detect a series of conflicting beliefs, they stop all interaction for those beliefs.

Example. Given the premise set $\Gamma = \{b_1(\neg p \vee \neg q),\ b_2 p,\ b_3 q,\ b_4(\neg p \vee r),\ b_5(\neg q \vee r)\}$, let $!b_i A$ abbreviate $b_i A \wedge \neg b_0 A$. The minimal *Dab*-consequences of Γ are:

$$Dab(\Delta_1) = Dab(!b_1(\neg p \vee \neg q),\ !b_2 p,\ !b_3 q)$$
$$Dab(\Delta_2) = Dab(!b_1(\neg p \vee \neg q),\ !b_2(p \vee q),\ !b_2(p \vee \neg q),\ !b_3(\neg p \vee q))$$
$$Dab(\Delta_3) = Dab(!b_1(\neg p \vee \neg q),\ !b_3(p \vee q),\ !b_2(p \vee \neg q),\ !b_3(\neg p \vee q))$$

$U(\Gamma) = \Delta_1 \cup \Delta_2 \cup \Delta_3$. Because all **DM**-models verify these *Dab*-formulas, there will be *reliable* models verifying e.g. $b_1(\neg p \vee \neg q) \wedge \neg b_0(\neg p \vee \neg q))$ and others verifying $b_2 p \wedge \neg b_0 p$. As a consequence neither $b_0(\neg p \vee \neg q)$ nor $b_0 p$ will be derivable according to this strategy. The interaction among the belief states $1-5$ amounts therefore to a merging only for consequences which are not among the formulas in $U(\Gamma)$.

Minimal Abnormality Strategy. A model is selected according to Minimal Abnormality if it verifies (in a set-theoretical sense) a minimal number of abnormalities. Hence, a **DM**-model M of Γ is a minimal abnormal model of Γ iff there is no **DM**-model M' of Γ such that $Ab(M') \subset Ab(M)$. This means that at each stage of the selection only one of the disjuncts of a *Dab*-consequence is considered true. The minimally abnormal models of Γ are those in respect to which the consequence relation of $\mathbf{ADM}^m$ is defined: for any $A \in \mathcal{B}$, $\Gamma \models_{\mathbf{ADM}^m} A$ iff A is verified by all minimal abnormal models of Γ. According to Minimal Abnormality, agents will try to make as many alliances as possible: these are to be intended as groupwise interaction among belief states. Any interaction provides one selection of abnormalities, whereas other merging failures are assumed to be false.

Example. Let us consider the premise set Γ of the previous example. In each of the minimal abnormal models $b_0(\neg p \vee r)$ and $b_0(\neg q \vee r)$ are verified. Moreover, in each of them $b_0((p \vee \neg q) \wedge p)$ is verified *or* $b_0((\neg p \vee \neg q) \wedge q)$ is *or* $b_0(p \wedge q)$ is. Hence, in all of them $b_0 r$ is true—note that this formula is false in some reliable models of Γ.

5 The Proof Theory of $\mathbf{ADM}^r$ and $\mathbf{ADM}^m$

Also the proof theory for ALs relies on the role of abnormalities. Typical of this dynamic proof theory is that a content may be formulated under condition of a *Dab*-formula, in which case it is derived unless one of the elements of the condition is true, i.e. provided these elements are false on the premises. In this way the dynamic proof theory of $\mathbf{ADM}^r$ and $\mathbf{ADM}^m$

establishes when a content should be withdrawn. The more insight is gained in the premise set and its consequences, the more contents are finally derived from it, i.e. derived and not later rejected. The standard deduction rules include: a rule to add premises on the empty conditon, an unconditional rule and a conditional one. At any stage of the proof, *Dab*-formulas may be derived on the empty condition; they are minimal if no proper subformula occurs in the proof on the empty condition. Given all the minimal *Dab*-formulas of a set Γ ($Dab(\Delta_1), Dab(\Delta_2), \ldots, Dab(\Delta_n)$) at stage s, the set of unreliable formulas with respect to Γ at that stage is given by their union $U_s(\Gamma) = \Delta_1 \cup \Delta_2 \cup \ldots \cup \Delta_n$. The two strategies provide the marking definitions to withdraw derived beliefs.

For the Reliability Strategy, a line i is marked at stage s iff Δ is the condition at line i and $\Delta \cap U_s(\Gamma) \neq \emptyset$, i.e. the condition of that line is an unreliable formula. The marking definition for the Minimal Abnormality Strategy requires that a proper minimal choice set $\Phi_s(\Gamma)$ be defined from the set of the minimal *Dab*-formulas at stage s of the premises Γ. Where $Dab(\Delta_1), \ldots, Dab(\Delta_n)$ are the minimal *Dab*-formulas derived on condition $\emptyset$ at stage s from Γ, $\Phi_s(\Gamma)$ is the set of minimal choice sets of $\{\Delta_1, \ldots, \Delta_n\}$. A line i in which A is derived on condition Δ is marked at stage s iff (i) there is no $\phi \in \Phi_s(\Gamma)$ such that $\phi \cap \Delta = \emptyset$, or (ii) for some $\phi \in \Phi_s(\Gamma)$ there is no line at which A is derived on a condition Θ for which $\phi \cap \Theta = \emptyset$. This means that the condition is part of the choice set of *Dab*-consequences of the premise set and at no later stage the same content is derived under another condition which is not a member of that choice set. Doxastic formulas derived from a set of premises Γ at a stage of the proof are derived at a line that is unmarked at that stage (according to one of the strategies). A doxastic formula is *finally derived* on a line i of a proof at stage s iff it is derived on a line i which is not marked at stage s, and any extension of the proof in which line i is marked, may be further extended so that it is unmarked. The definitions of merging according to the strategies are the following:

ADM^r In view of a premise set Γ, a doxastic formula $b_i\phi$ is mergeable according to the Reliability Strategy if there is an **ADM**^{r}-proof from Γ in which $b_0\phi$ is finally derived.

ADM^m In view of a premise set Γ, a doxastic formula $b_i\phi$ is mergeable according to the Minimal Abnormality Strategy if there is an **ADM**^{m}-proof from Γ in which $b_0\phi$ is finally derived.

Further extensions of **ADM** may be given in terms of the dynamics of preference, dynamic belief revision (i.e. of an update operator) and second degree doxastic operators.

BIBLIOGRAPHY

[AGM85] C.E Alchourrón, P. Gärdernfors, and D. Makinson. On the logic of theory change: partial meet contraction and revision functions. *Journal of Symbolic Logic,*, 50:510–530, 1985.

[All06] P. Allo. Local information and adaptive consequence. *Logique & Analyse*, 196:461–488, 2006.

[Bat01] D. Batens. A general characterization of adaptive logics. *Logique & Analyse*, 173-175:45–68, 2001.

[Bat07] D. Batens. A universal logic approach to adaptive logics. *Logica universalis*, 1:221–242, 2007.

[KPP02] S. Koneczny and R. Pino-Pérez. Merging information under constraints: A logical framework. *Journal of Logic and Computation*, 12(5):773–808, 2002.

[LM99] J. Lin and A.O. Mendelzon. Knowledge base merging by majority. In *Dynamic Worlds: from the Frame Problem to Knowledge Management*. Kluwer, 1999.

[LS98] P. Liberatore and M. Schaerf. Arbitration (or how to merge knowledge bases). *IEEE Transactions on Knowledge and Data Engineering*, pages 76–90, 1998.

[LS07] H. Leitgeb and K. Segerberg. Dynamic doxastic logic: why, how and where to? *Synthese*, 155:167–190, 2007.

[vB96] J. van Benthem. *Exploring Logical Dynamics*. CSLI & Folli Publications, 1996.

[vD05] H.P. van Ditmarsch. Prolegomena to dynamic logic for belief revision. *Synthese*, 147:229–275, 2005.