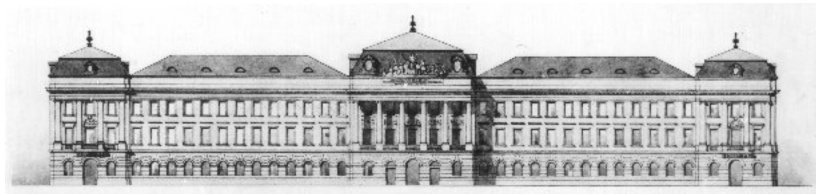
I N F S Y S

R E S E A R C H

R E P O R T



INSTITUT FÜR INFORMATIONSSYSTEME

ARBEITSBEREICH WISSENSBASIERTE SYSTEME

# FINDING EXPLANATIONS OF INCONSISTENCY IN MULTI-CONTEXT SYSTEMS

THOMAS EITER MICHAEL FINK PETER SCHÜLLER
ANTONIUS WEINZIERL

INFSYS RESEARCH REPORT 1843-12-09

DECEMBER 2012

Institut für Informationssysteme
AB Wissensbasierte Systeme
Technische Universität Wien
Favoritenstrassße 9-11
A-1040 Wien, Austria
Tel: +43-1-58801-18405
Fax: +43-1-58801-18493
sek@kr.tuwien.ac.at
www.kr.tuwien.ac.at

INFSYS RESEARCH REPORT

INFSYS RESEARCH REPORT 1843-12-09, DECEMBER 2012

# FINDING EXPLANATIONS OF INCONSISTENCY IN MULTI-CONTEXT SYSTEMS

Thomas Eiter[1] Michael Fink[1] Peter Schüller [2] Antonius Weinzierl[1]

**Abstract.** Interlinking knowledge sources to enable information exchange is a basic means to build enriched knowledge-based systems, which gains importance with the spread of the Internet. Inconsistency, however, arises easily in such systems, which is not least due to their heterogeneity, but also due to their independent design. This makes developing methods for consistency management of such systems a pressing issue. An important aspect is that in many relevant cases, the information at individual sources may not be amenable to change in order to resolve inconsistency, like in case of autonomous management of the sources. We thus aim at analyzing inconsistency of a system by means of the interlinking of sources and changes thereof. More concretely, we consider the powerful framework of Multi-Context Systems, in which decentralized and heterogeneous system parts interact via (possibly nonmonotonic) bridge rules for information exchange. Nonmonotonicity and potential cyclic dependencies pose additional challenges that call for suitable methods of inconsistency analysis. We thus provide two approaches for explaining inconsistency, which both characterize inconsistency in terms of bridge rules, but in different ways: by pointing out rules which need to be altered for restoring consistency, and by finding combinations of rules which cause inconsistency. We show duality and modularity properties of these notions, give precise complexity characterizations, and provide algorithms for their computation, which have been implemented in a prototype, by means of so-called HEX-programs. Our results provide a basis for inconsistency management in heterogeneous knowledge systems which, different from and orthogonal to other works, explicitly addresses the knowledge interlinks in order to restore consistency.

[1]Institute of Information Systems, Knowledge-Based Systems Group, TU Vienna, Favoritenstraße 9-11, A-1040 Vienna, Austria. Email: (eiter | fink | weinzierl)@kr.tuwien.ac.at

[2]Cognitive Robotics Laboratory, Faculty of Engineering and Natural Sciences, Sabancı University, Orhanli, Tuzla - 34956, Istanbul, Turkey. Email: peterschueller@sabanciuniv.edu

**Acknowledgements**: This work was supported by the Vienna Science and Technology Fund (WWTF) under grant ICT 08-020.

This article is a revised and significantly extended version of the preliminary paper with the same title presented at the Twelfth International Conference on Principles of Knowledge Representation and Reasoning (KR 2010), Toronto, Canada, May 9-13, 2010.

# Contents

# 1 Introduction

In recent years, there has been increasing interest in interlinking information, driven by—and reflected in—the development of the World Wide Web. Respective progress and ever increasing demands not only regard the large number of available sources and the degree of interlinking, but also the quality of the information exchanged. From its initial conception as a means to link textual data, the Internet has evolved to a medium for interlinking, accessing, and exchanging also more structured information including relational and semi-structured data, and in the last years also semantically richer knowledge sources. Systems may thus be built in which individual information sources are connected, such that more informed and accurate answers can be given to specific user problems. Typically, these sources are expressed in different formalisms, and they are autonmously managed by third parties, such that a real integration is difficult.

Developing uniform, high-level formalisms to capture such systems has thus become an relevant issue in knowledge representation and reasoning (KRR). The Multi-Context System framework (see for example the papers [Ghidini and Giunchiglia, 2001], [Roelofsen and Serafini, 2005], [Brewka *et al.*, 2007], [Brewka and Eiter, 2007], [Bikakis and Antoniou, 2010], [Serafini and Homola, 2012]), which evolved from MultiLanguage systems [Giunchiglia and Serafini, 1994, Giunchiglia, 1993], is an expressive framework for this purpose. It is a powerful knowledge representation formalism for many application scenarios where heterogeneity and pointwise, inter-contextual information exchange are essential properties. Multi-context systems (MCS) as introduced by [Brewka and Eiter, 2007] consist of knowledge bases (in possibly heterogeneous and/or nonmonotonic logics) at nodes (called *contexts*) that formulate the exchange of information via so called bridge rules, such as

$$\begin{aligned}(c_1\colon ok(X\_Risk) \leftarrow (&c_2\colon insurance(X\_Risk),\\ &(c_2\colon low\_rate(X\_Risk), \mathbf{not}\ (c_3 : black\_list(X\_Risk));\end{aligned}$$

informally, it says that if $X\_Risk$ is a low rate insurance company according to knowledge source (context) $c_2$, and it is not known to be blacklisted according to context $c_3$, then context $c_1$ adds the fact $ok(X\_Risk)$ to its knowledge base.

MCS enable knowledge exchange at a general level, by interlinking possibly heterogenous formalisms like e.g. ontologies, databases, and logic programs. However, due to their decentralized nature, information exchange can have unforeseen effects, and in particular cause an MCS to be inconsistent. For example, consider a system for supporting health care decisions in a hospital, that comprises several components: (i) a database of laboratory test results; (ii) a patient record database; (iii) an ontology for disease classification; and (iv) an expert system suggesting suitable treatments for patients. Modeled as an MCS, each component is a context and the information flow between them is specified by suitable bridge rules (cf. Example 5 for details); thanks to the latter, existing systems might be easily incorporated. Suppose the expert system concludes that a patient must be given a special drug, but the patient database states that she is allergic to that drug, thus counter-indicating its use. The whole system gets inconsistent if such special cases were not anticipated when contexts and bridge rules were modeled, rendering it useless.

In real world applications, system complexity tends to increase, in terms of both contexts and interconnectivity. Extensive testing considering all possible states of a system often is infeasible, especially if legacy information systems are employed as contexts in an MCS. Therefore, and specifically due to the heterogeneity of individual system components that are linked together, inconsistency arises easily and methods for handling inconsistency are a pressing issue. Our work aims at *analyzing* inconsistencies in MCS, in order to understand where and why such inconsistencies occur, and how they might be removed. It thus provides a basis for the specification of concrete strategies to handle inconsistencies and to extend systems with inconsistency management mechanisms (in addition to some basic operations that can be obtained directly from our approach).

Various approaches to cope with inconsistent information have been developed for different KRR formalisms (see Section 7 for works closely related to ours and a more general overview). Specifically in traditional data integration scenarios, inconsistency problems also surface naturally and methods for consistency restoring or maintenance have been studied extensively. Compared to that, however, our work *innovates conceptually and differs considerably* beyond differences in the settings and formalisms considered: The important point is that we *focus on the exchange of information, its interlinking*, i.e., on *adjusting bridge rules* instead of modifying the data in the contexts. In data integration terms, we thus consider modifications of the mapping as a potential for resolving inconsistency rather than repairing or cleaning the data. While the importance of maintaining and repairing mappings has been recognized in database integration [Doan and Halevy, 2005], progress on this has been on a slow pace.

The motivation for addressing inconsistency in MCS on the level of bridge rules stems from the fact that an MCS models usually a loose integration scenario with autonomous sources (e.g., if companies link their business logics), where changing contexts or their data to restore consistency may not be an option. Compared to data integration settings that globally materialize (at least virtually) the data of the individual sources, it may not be possible to modify (or even access all the) data that is internal to one of the knowledge bases employed in an MCS; in particular, this applies to cases of cyclic information exchange. Therefore, we identify bridge rules as the source of inconsistency, and their modification as a possibility of counteracting. On the one hand, under the reasonable assumption that every context is consistent if bridge rules are disregarded, we can fully capture reasons of inconsistency in terms of bridge rules. On the other hand, negation and potential cyclic dependencies (as opposed to acyclic mappings in data integration) render the task of characterizing and analyzing inconsistency non-trivial.

Starting from this, our contributions are summarized as follows:

(1) In the spirit of debugging approaches used in the nonmonotonic reasoning community, especially in logic programming [Inoue and Sakama, 1995], [Syrjänen, 2006], [Brain *et al.*, 2007], [Pereira *et al.*, 1993a], [Lloyd, 1987], we introduce two notions of explaining inconsistency in MCS: a *consistency-based* notion, called *diagnosis*, which characterizes inconsistency in terms of modified sets of bridge rules that are consistent, and an *entailment-based* notion, called *explanation*, which captures bridge rules that cause inconsistency in a given system. Potential nonmonotonicity makes intuitive and sound notions challenging; that our notions have appealing properties may count as evidence for their suitability. Further refinements and restrictions of our notions are investigated; we show, however, that these can be expressed in terms of our basic notions.

(2) We establish useful properties of our notions. First, conversion and duality results between diagnoses and explanations show that, while representing different analytic properties, they identify the same overall set of bridge rules as relevant for inconsistency. This in fact generalizes a similar classic result by Reiter [Reiter, 1987], who characterized the consistency-based diagnoses of system description in classical (monotonic) logic in terms of conflict sets. Furthermore, we establish modularity properties in the spirit of Splitting Sets [Lifschitz and Turner, 1994], which allow for an incremental computation of diagnoses and explanations, taking the MCS topology into account.

(3) We sharply characterize the computational complexity of identifying explanations, under varying assumptions on the complexity of reasoning in contexts (note that by the underlying assumption, consistency-based explanations always exist). It turns out that this problem has for a range of context complexities no (or only mildly) higher complexity than the contexts themselves. As a consequence, computing explanations is in some cases not harder than consistency checking.

(4) Finally, we consider how consistency- and entailment-based explanations can be computed. The is end, we resort here to HEX-programs, which are a generalization of Answer Set Programming (ASP) by so-called external atoms that provide access to external sources of computation; an experimental prototype has been implemented, while more advanced implementatios are underway.

Our results provide a basis for building enhanced MCS systems which are capable of analyzing and reasoning about emerging inconsistencies. Rather than automatically resolving inconsistency, as e.g. in [Bikakis and Antoniou, 2008], we envisage a (semi-)automatic approach with user support for locating and tracking parts that cause inconsistency. Indeed, user invention may be indispensable as often no automatic solution is feasible; in our healthcare example, for instance, giving the special drug would resolve the inconsistency, but this should only be done after approval by a medical doctor.

**Structure.** The remainder of this article is organized as follows. Section 2 provides necessary preliminaries on the MCS framework and introduces a running example. In Section 3, diagnoses and explanations are introduced, while Section 4 contains conversion results and modularity properties that depend on the interlinking of MCS. Section 5 contains a detailed analysis of the computational complexity of identifying both diagnoses and explanations, and in Section 6 we elaborate on how they can be computed using HEX programs. A comprehensive discussion of related work is given in Section 7, followed by concluding remarks and directions for future work in Section 8. Further material is provided in the appendix; some of our examples are detailed more in-depth in A, while B contains proofs for all results.

# 2 Preliminaries

In this section, we recall nonmonotonic MCS from [Brewka and Eiter, 2007]. Further background is given in [Brewka *et al.*, 2011a], [Eiter *et al.*, 2009], which discuss extensions of MCS, compare them to other related formalisms, and survey computational issues.

Loosely speaking, a nonmonotonic MCS consists of *contexts*, each composed of a knowledge base with an underlying abstract *logic*, and a set of *bridge rules* which control the information flow between contexts. The MCS framework uses a minimalistic, abstract model of logics, which consists of possible sets of formulas, possible sets of beliefs, and a satisfiability relation.

**Definition 1.** *A logic* $L = (\mathbf{KB}_L, \mathbf{BS}_L, \mathbf{ACC}_L)$ *consists, in an abstract view, of the following components:*

- $\mathbf{KB}_L$ *is the set of well-formed knowledge bases of* $L$*. We assume each element of* $\mathbf{KB}_L$ *is a set (of "formulas").*
- $\mathbf{BS}_L$ *is the set of possible belief sets, where the elements of a belief set are statements that possibly hold, given a knowledge base.*
- $\mathbf{ACC}_L : \mathbf{KB}_L \to 2^{\mathbf{BS}_L}$ *is a function describing the "semantics" of the logic by assigning to each knowledge base a set of acceptable belief sets.*

Intuitively, a belief set is a set of statements (beliefs) that a reasoner may jointly hold, and $\mathbf{ACC}_i(kb)$ singles out, given a knowlede base $kb$, the belief sets that are acceptable according to some reasoning rationale; to accommodate nonmonotonic formalisms where knowledge basy may have multiple such acceptable belief sets (e.g., answer sets of logic program; extensions of a default theory; expansions of an autoepistemic theory), $\mathbf{ACC}_i(kb)$ is designed as a multi-valued function.

This abstract notion of a purely functional "logic" captures many monotonic and nonmonotonic logics, e.g., classical logic, description logics, modal, default, and autoepistemic logics, circumscription, and logic programs under answer set semantics.

The following examples introduce logics used in formalizing our running example, and they shall illustrate how this abstraction captures some well-known knowledge-representation formalisms.

**Example 1.** *To capture classical (propositional) logic over a set $At$ of propositional atoms, we may define:*

- $\mathbf{KB}^c = 2^\Sigma$ *is the set of all subsets of* $\Sigma$*, where* $\Sigma$ *is the set of well-formed formulas over* $At$ *built using the connectives* $\wedge, \vee, \neg, \rightarrow$*;*
- $\mathbf{BS}^c = 2^\Sigma$*, i.e., each set of formulas is a possible belief set; and*
- $\mathbf{ACC}^c$ *returns for each set* $kb \in \mathbf{KB}^c$ *of well-formed formulas a singleton set that contains the set of formulas entailed by* $kb$*; if* $\models_c$ *denotes classical entailment, then* $\mathbf{ACC}^c(kb) = \{\{F \in \Sigma \mid kb \models_c F\}\}$.

*The resulting logic* $L^c_\Sigma = (\mathbf{KB}^c, \mathbf{BS}^c, \mathbf{ACC}^c)$ *captures entailment in classical logics.*

*In practice, the formulas in knowledge bases and belief sets might be restricted to particular forms, e.g., to literals; we denote the respective logic by* $L^{pl}_\Sigma = (\mathbf{KB}^{pl}, \mathbf{BS}^{pl}, \mathbf{ACC}^{pl})$*. Note that*

$$\mathbf{ACC}^{pl}(kb) = \{\{A \in At \mid kb \models_c A\} \cup \{\neg A \in At \mid kb \models_c \neg A\}\}\,.$$

**Example 2.** *For our running example we employ two contexts,* $C_1$ *and* $C_2$*, using the respective signatures* $\Sigma_1 = \{\mathit{allergy\_strong\_ab}\}$ *and* $\Sigma_2 = \{\mathit{blood\_marker}, \mathit{xray\_pneumonia}\}$*, and logics* $L^{pl}_{\Sigma_1}$ *and* $L^{pl}_{\Sigma_2}$*, respectively. Their knowledge bases are as follows:*

$$kb_1 = \{\mathit{allergy\_strong\_ab}\},$$
$$kb_2 = \{\neg \mathit{blood\_marker}, \mathit{xray\_pneumonia}\},$$

*Those knowledge bases provide information that the patient is allergic to strong antibiotics (*$kb_1$*), respectively that a certain blood marker is not present and that pneumonia was detected in an X-ray examination (*$kb_2$*).*

*The corresponding semantics is given by* $\mathbf{ACC}(kb_1) = \big\{\{\mathit{allergy\_strong\_ab}\}\big\}$ *for* $C_1$*, and* $\mathbf{ACC}(kb_2) = \big\{\{\neg \mathit{blood\_marker}, \mathit{xray\_pneumonia}\}\big\}$ *for* $C_2$*.*

**Example 3.** *For ontologies with syntax and semantics of the description logic* $\mathcal{ALC}$ *(cf. [Baader* et al.*, 2003]), we use the abstract logic* $L_\mathcal{A}$ *obtained similar as* $L^{pl}_\Sigma$ *above for propositional logic (for details see Example 29 in Appendix A). Intuitively,* $L_\mathcal{A}$ *captures T-Box axioms and A-Box axioms of the description logic* $\mathcal{ALC}$*. An* $L_\mathcal{A}$*-knowledge base contains both, A-Box and T-Box axioms. An accepted belief set of such a knowledge base is the set of assertions that follow from it.*

*For a running example, we use an ontology about diseases, given by context* $C_3$ *using* $L_\mathcal{A}$*. Its knowledge base,* $kb_3$*, consists of two axioms, where the first states that pneumonia is a bacterial disease and the second states that pneumonia which has an associated blood-marker implies an atypical pneumonia (that is a severe form of pneumonia). The corresponding knowledge base is:*

$$\begin{aligned} kb_3 = \{&\mathit{Pneumonia} \sqsubseteq \mathit{BacterialDisease},\\ &\mathit{Pneumonia} \sqcap \exists \mathit{has\_marker}.\top \sqsubseteq \mathit{AtypPneumonia}\}. \end{aligned}$$

*As* $kb_3$ *is satisfiable and contains only terminological knowledge, no assertions follow from this knowledge base, thus* $\mathbf{ACC}(kb_3) = \{\emptyset\}$*. Adding the assertion that* $d$ *is pneumonia results in the conclusion that* $d$ *also is a bacterial disease, i.e.,* $\mathbf{ACC}\big(kb_3 \cup \{d : \mathit{Pneumonia}\}\big) = \big\{\{d : \mathit{Pneumonia}, d : \mathit{BacterialDisease}\}\big\}$.

**Example 4.** *For normal disjunctive logic programs under answer set semantics over a non-ground signature $\Sigma$ (cf. [Przymusinski, 1991] and [Faber* et al.*, 2004]), we use the abstract logic $L^{asp}_{\Sigma}$, which is detailed in the Appendix in Example 30. We employ $L^{asp}_{\Sigma}$ for a context, $C_4$, suggesting proper treatments where $\Sigma = \{give\_strong, give\_weak, need\_ab, allow\_strong\_ab, give\_nothing\}$. The knowledge base for $C_4$ is:*

$$\begin{aligned} kb_4 = \{ & give\_strong \vee give\_weak \leftarrow need\_ab. \\ & give\_strong \leftarrow need\_strong. \\ & \bot \leftarrow give\_strong, not\ allow\_strong\_ab. \\ & give\_nothing \leftarrow not\ need\_ab, not\ need\_strong. \}. \end{aligned}$$

*$C_4$ suggests a treatment which is either a strong antibiotics, a weak antibiotics, or no medication at all. Without further information, $kb_4$ thus concludes that nothing is required, i.e., $\mathbf{ACC}(kb_4) = \big\{\{give\_nothing\}\big\}$. If $need\_ab$ is added, however, $kb_4$ results in two answer sets, i.e., $\mathbf{ACC}\big(kb_4 \cup \{need\_ab.\}\big) = \{A_1, A_2\}$ where $A_1 = \{give\_strong, need\_ab\}$ and $A_2 = \{give\_weak, need\_ab\}$.*

A *bridge rule* can add information to a context, depending on the belief sets which are accepted at other contexts. Let $L = (L_1, \ldots, L_n)$ be a sequence of logics. A $L$-bridge rule $r$ over $L$ is of the form

$$(k : s) \leftarrow (c_1 : p_1), \ldots, (c_j : p_j), \mathbf{not}\ (c_{j+1} : p_{j+1}), \ldots, \mathbf{not}\ (c_m : p_m). \quad (1)$$

where $1 \leq c_i \leq n$, $p_i$ is an element of some belief set of $L_{c_i}$, and $s$ is a knowledge-base formula of $L_k$ (i.e. $s \in \bigcup \mathbf{KB}_{L_k}$). We denote by $h_b(r)$ the formula $s$ in the head of $r$ and by $h_c(r)$ the context $k$ where $r$ belongs to. The full head of $r$ is denoted by $hd(r) = (k : s)$. The literals in the body of $r$ are referred to by $body(r)$, $body^+(r)$, $body^-(r)$ which denotes the set $\{(c_1 : p_1), \ldots, (c_m : p_m)\}$, $\{(c_1 : p_1), \ldots, (c_j : p_j)\}$, $\{(c_{j+1} : p_{j+1}), \ldots, (c_m : p_m)\}$, respectively. Furthermore, $b_c(r)$ denotes the set of contexts referenced in $r$'s body, i.e., $b_c(r) = \{c_i \mid (c_i : p_i) \in body(r)\}$. For technical use later, we denote by $cf(r)$ the *condition-free* bridge rule stemming from $r$ by removing all elements in its body, i.e., $cf(r)$ is $(k : s) \leftarrow .$ and for any set of bridge rules $R$, we let $cf(R) = \bigcup_{r \in R} cf(r)$.

**Definition 2.** *A multi-context system $M = (C_1, \ldots, C_n)$ is a collection of contexts $C_i = (L_i, kb_i, br_i)$, $1 \leq i \leq n$, where $L_i = (\mathbf{KB}_i, \mathbf{BS}_i, \mathbf{ACC}_i)$ is a logic as above, $kb_i \in \mathbf{KB}_i$ a knowledge base, and $br_i$ is a set of $L$-bridge rules over $L = (L_1, \ldots, L_n)$. Furthermore, for each $H \subseteq \{h_b(r) \mid r \in br_i\}$ it holds that $kb_i \cup H \in \mathbf{KB}_{L_i}$, i.e., adding bridge rule heads to a knowledge base yields a legal knowledge base.*

By $br(M) = \bigcup_{i=1}^{n} br_i$ and $c(M) = \{C_1, \ldots, C_n\}$ we denote the set of all bridge rules, resp. the set of all contexts of $M$. We write $br_i(M)$ to denote the set of bridge rules of context $i$ of $M$, i.e., $br_i(M) = \{r \in br(M) \mid h_c(r) = i\}$.

In the following, we formally introduce our running example.

**Example 5.** *Consider an MCS $M$ embodying a health care decision support system that contains the following contexts: a patient history database ($C_1$), a blood and X-Ray analysis database ($C_2$), a disease ontology ($C_3$), and an expert system ($C_4$) which suggests proper treatments. The corresponding abstract logics and knowledge bases are those in Examples 2, 3, and 4.*

*The bridge rules are as follows:*

$$\begin{aligned} r_1\!: &\quad (3 : d{:}Pneumonia) \leftarrow (2 : xray\_pneumonia). \\ r_2\!: &\quad (3 : (d, m1){:}has\_marker) \leftarrow (2 : blood\_marker). \\ r_3\!: &\quad (4 : need\_ab.) \leftarrow (3 : d{:}BacterialDisease). \\ r_4\!: &\quad (4 : need\_strong.) \leftarrow (3 : d{:}AtypPneumonia). \\ r_5\!: &\quad (4 : allow\_strong\_ab.) \leftarrow \mathbf{not}\ (1 : allergy\_strong\_ab). \end{aligned}$$

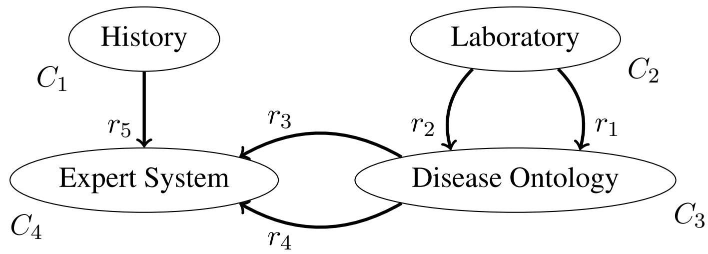



Figure 1: Knowledge bases and bridge rules of the hospital MCS.

*Rules $r_1$ and $r_2$ provide input for disease classification to the ontology; they assert facts about a new individual '$p$' corresponding to the patient. Rules $r_3$ and $r_4$ link disease information with medication requirements, while $r_5$ relates acceptance of strong antibiotics with an allergy check on the patient database.*

*A layout of the information exchange in this MCS, is depicted in Figure 1, where each bridge rule $r \in \{r_1, \ldots, r_5\}$ with $h_c(r) = j$ and $i \in b_c(r)$ is depicted as an arrow from $C_i$ to $C_j$.*

The semantics of MCS is defined over locally accepted belief sets, taking the bridge rules into account as follows. A *belief state* of an MCS $M = (C_1, \ldots, C_n)$ is a sequence $S = (S_1, \ldots, S_n)$ of belief sets $S_i \in \mathbf{BS}_i$, $1 \leq i \leq n$. A bridge rule $r$ of form (1) is *applicable* in $S$, denoted $S \models\!\!\leadsto r$, iff for all $(c : p) \in \mathit{body}^+(r)$ it holds that $p \in S_c$, and for all $(c : p) \in \mathit{body}^-(r)$ it holds that $p \notin S_c$. For a set $R$ of bridge rules, $\mathit{app}(R, S) = \{r \in R \mid S \models\!\!\leadsto r\}$, denotes the set of applicable bridge rules.

Equilibrium semantics selects a belief state $S$ of an MCS $M$ as acceptable, if each context $C_i$ takes the heads of all bridge rules that are applicable in $S$ into account, and accepts $S_i$.

**Definition 3.** *A belief state $S = (S_1, \ldots, S_n)$ of $M$ is an equilibrium iff for all $1 \leq i \leq n$, $S_i \in \mathbf{ACC}_i(kb_i \cup \{h_b(r) \mid r \in \mathit{app}(br_i, S)\})$.*

**Example 6.** *In our example, $M$ has a single equilibrium $S = (S_1, S_2, S_3, S_4)$ where*

$$\begin{aligned}
S_1 =& \{\mathit{allergy_strong_ab}\},\\
S_2 =& \{\neg \mathit{blood_marker}, \mathit{xray_pneumonia}\},\\
S_3 =& \{d : \mathit{Pneumonia}, d : \mathit{BacterialDisease}\},\\
S_4 =& \{\mathit{need_ab}, \mathit{give_weak}\}).
\end{aligned}$$

*The only rules applicable in $S$ are $r_1$ and $r_3$, because $\mathit{app}(br_1(M), S) = \mathit{app}(br_2(M), S) = \emptyset$, $\mathit{app}(br_3(M), S) = \{r_1\}$, and $\mathit{app}(br_4(M), S) = \{r_3\}$. Note that if we replace $S_4$ with the set $\{\mathit{need\_ab}, \mathit{give\_strong}, \mathit{allow\_strong\_ab}\}$, then the resulting belief state is not an equilibrium: $C_4$ uses answer set semantics, therefore $\mathit{allow\_strong\_ab}$ cannot be part of $S_4$ unless it is added by a bridge rule. The only bridge rule with this head is $r_5$, and its applicability is blocked by the presence of $\mathit{allergy\_strong\_ab}$ in $kb_1$ and in $S_1$.*

## 3 Diagnoses and Explanations for Inconsistency

*Inconsistency* in an MCS is the lack of an equilibrium. As the combination and interaction of heterogeneous, possibly autonomous, systems can easily have unforeseen and intricate effects, inconsistency is a major problem in MCS. To provide support for restoring consistency, we seek to understand and give reasons for inconsistency.

**Example 7.** *As a running example, we consider a slightly modified version of Example 5, where the blood serum analysis shows presence of the blood marker:*

$$kb_2 = \{blood\_marker, xray\_pneumonia\}.$$

*This MCS is inconsistent since $r_2$ and $r_4$ become applicable, which in turn requires strong antibiotics. This is in conflict with the patient's allergy. Note that applicability of $r_5$ would resolve this inconsistency by activating $allow\_strong\_ab$. However, presence of $allergy\_strong\_ab$ in $S_1$ together with the body atom '$\mathbf{not}\ (1 : allergy\_strong\_ab)$' in $r_5$ prevents the applicability of $r_5$ (due to negation as failure).*

We will use the following notation. Given an MCS $M$ and a set $R$ of bridge rules (that are compatible with $M$), we denote by $M[R]$ the MCS obtained from $M$ by replacing its set of bridge rules $br(M)$ with $R$; e.g., $M[br(M)] = M$ and $M[\emptyset]$ is $M$ with no bridge rules. By $M \models \bot$ we denote that $M$ has no equilibrium, i.e., is inconsistent, and by $M \not\models \bot$ the opposite.

In the following, we consider two possibilities for explaining inconsistency in MCS: first, a consistency-based formulation, which identifies a part of the bridge rules which need to be changed to restore consistency. Second, an entailment-based formulation, which identifies a part of the bridge rules which is required to make the MCS inconsistent. Following common terminology, we call the first formulation *diagnosis* (cf. [Reiter, 1987]) and the second *inconsistency explanation*.

### 3.1 Diagnoses

As well-known, adding knowledge in nonmonotonic reasoning can both cause and prevent inconsistency; the same is true for removing knowledge.

For our consistency-based explanation of inconsistency, we therefore consider pairs of sets of bridge rules, such that if we deactivate the rules in the first set, and add the rules in the second set in unconditional form, the MCS becomes consistent (i.e., admits an equilibrium). Adding rules unconditionally is the most severe form of modification of a rule's body, but as we later see, this notion also allows to capture more fine-grained forms of modification.

**Definition 4.** *Given an MCS $M$, a* diagnosis *of $M$ is a pair $(D_1, D_2)$, $D_1, D_2 \subseteq br(M)$, such that $M[br(M) \setminus D_1 \cup cf(D_2)] \not\models \bot$. By notation, $D^{\pm}(M)$ is the set of all diagnoses.*

To obtain a more relevant set of diagnoses, by Occam's razor we prefer subset-minimal diagnoses, where for pairs $A = (A_1, A_2)$ and $B = (B_1, B_2)$ of sets, the pointwise subset relation $A \subseteq B$ holds iff $A_1 \subseteq B_1$ and $A_2 \subseteq B_2$.

**Definition 5.** *Given an MCS $M$, $D^{\pm}_m(M)$ is the set of all pointwise subset-minimal diagnoses of an MCS $M$, i.e.,*

$$D^{\pm}_m(M) = \{D \in D^{\pm}(M) \mid \forall D' \in D^{\pm}(M) : D' \subseteq D \Rightarrow D \subseteq D'\}.$$

**Example 8.** *In our running example, we obtain*

$$D^{\pm}_m(M) = \big\{(\{r_1\}, \emptyset), (\{r_2\}, \emptyset), (\{r_4\}, \emptyset), (\emptyset, \{r_5\})\big\}.$$

*Accordingly, deactivating $r_1$, or $r_2$, or $r_4$, or adding $r_5$ unconditionally, will result in a consistent MCS.*

*In more detail, we find: diagnosis $(\{r_1\}, \emptyset)$ removes bridge rule $r_1$. This way we ignore the X-Ray finding and obtain the following equilibrium:*

$$\begin{aligned} EQ_1 = (&\{allergy\_strong\_ab\}, \{blood\_marker, xray\_pneumonia\},\\ &\{(d, m1) : has\_marker\}, \{give\_nothing\}). \end{aligned}$$

*It represents that we do not treat the patient since no illness is detected in it.*

*Diagnosis* $(\{r_4\}, \emptyset)$ *removes bridge rule* $r_4$. *This ignores the information that treating the illness requires the strong antibiotics. We obtain the following equilibrium:*

$$\begin{aligned} EQ_2 = (&\{allergy\_strong\_ab\}, \{blood\_marker, xray\_pneumonia\},\\ &\{(d, m1){:}has\_marker, d{:}Pneumonia, d{:}BacterialDisease,\\ &\ d{:}AtypPneumonia\}, \{need\_ab, give\_weak\}). \end{aligned}$$

*Diagnosis* $(\emptyset, \{r_5\})$ *adds an unconditional copy of bridge rule* $r_5$*, which forces strong antibiotics to be allowed as a treatment. The modified system has the following equilibrium:*

$$\begin{aligned} EQ_3 = (&\{allergy\_strong\_ab\}, \{blood\_marker, xray\_pneumonia\},\\ &\{(d, m1){:}has\_marker, d{:}Pneumonia, d{:}BacterialDisease,\\ &\ d{:}AtypPneumonia\}, \{need\_ab, need\_strong, allow\_strong\_ab,\\ &\ give\_strong\}). \end{aligned}$$

*Any or none of the above possibilities might be the right choice: such decisions ought to be taken by a domain specialist (e.g., a doctor) and cannot be done automatically. Therefore analysis of inconsistency is important to identify reasons for it.*

Preference on diagnoses can be defined in general, relying on some notion of plausibility (see e.g., for abduction [Bylander *et al.*, 1991b]). This is, however, beyond the scope of this work, as we investigate basic notions of inconsistency here.

### 3.2 Explanations.

In the spirit of abductive reasoning, we also propose an entailment-based notion of explaining inconsistency. An *inconsistency explanation* (in short, an *explanation*) is a pair of sets of bridge rules, whose presence or, expected, absence entails a relevant inconsistency in the given MCS.

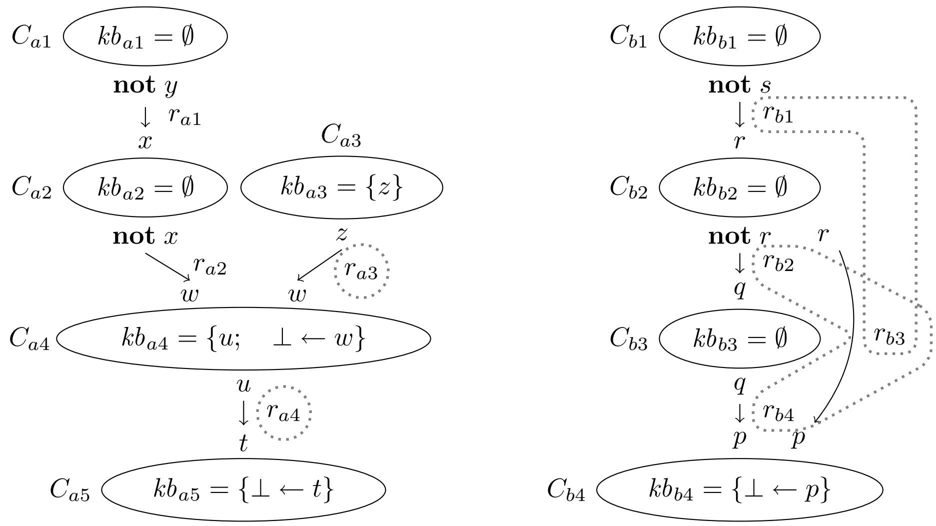
**Definition 6.** *Given an MCS* $M$*, an* inconsistency explanation *of* $M$ *is a pair* $(E_1, E_2)$ *of sets* $E_1, E_2 \subseteq br(M)$ *of bridge rules, such that for all* $(R_1, R_2)$ *where* $E_1 \subseteq R_1 \subseteq br(M)$ *and* $R_2 \subseteq br(M) \setminus E_2$*, it holds that* $M[R_1 \cup cf(R_2)] \models \bot$. *By* $E^{\pm}(M)$ *we denote the set of all inconsistency explanations of* $M$*, and by* $E^{\pm}_m(M)$ *the set of all pointwise subset-minimal ones.*

The intuition about $E_1$ is as follows: bridge rules in $E_1$ are crucial to create an inconsistency in $M$ (i.e., $M[E_1] \models \bot$), and this inconsistency is relevant for $M$ in the sense that adding some bridge rules from $br(M)$ to $M[E_1]$ never yields a consistent system.

This condition of relevancy is necessary for non-monotonic reasoning systems; for example the program $P = \{a \leftarrow not\, a.\}$ is inconsistent under the answer set semantics, but its superset $P' = \{a \leftarrow not\, a.\ \ a.\}$ is consistent. The inconsistency of $P$ does not matter for $P'$. In terms of MCS, a set of bridge rules may create an inconsistency in $M$, but this inconsistency is irrelevant, as it does not occur if more or all bridge rules are present.

The intuition about $E_2$ regards inconsistency wrt. the application of bridge rules: $M[E_1]$ cannot be made consistent unless at least one bridge rule from $E_2$ fires.

In summary, bridge rules $E_1$ create a relevant inconsistency, and at least one bridge rule in $E_2$ must applied in unconditional form to repair that inconsistency.




(a) Example MCS $M_a$

(b) Example MCS $M_b$

Figure 2: Example MCS topologies for illustrating properties and the usefulness of inconsistency explanations. Dotted areas indicate individual inconsistency explanations.

**Example 9.** *In our running example, we have one minimal inconsistency explanation, namely* $(\{r_1, r_2, r_4\}, \{r_5\})$. *To trigger the only possible inconsistency, which is in* $C_4$*, we need to import* $\mathit{need\_strong}$ *(using* $r_4$*) and we must not import* $\mathit{allow\_strong\_ab}$ *(using* $r_5$*). Furthermore,* $r_4$ *can only fire if* $C_3$ *accepts* $\mathit{AtypPneumonia}(p)$*, which is only possible if* $r_1$ *and* $r_2$ *fire. Therefore,* $r_1$*,* $r_2$*, and* $r_4$ must *be present to get inconsistency, and the head of* $r_5$ must not *be present.*

From Definition 6 the following property follows immediately.

**Proposition 1.** *Given an explanation* $E$ *of an MCS* $M$*, every* $E'$ *such that* $E \subseteq E' \subseteq br(M) \times br(M)$ *is also an explanation.*

We now give further examples of inconsistency explanations and their properties.

**Example 10.** *Consider a modification of our running medical example, where bridge rules are added for the administration of anti-allergenics. Bridge rule* $r_6$ *encodes that an allergy blocking (anti-allergenic) medication is given, if there strong antibiotics is needed, the patient is allergic to it, and nothing was done to block the allergic reaction;* $r_7$ *encodes that the patient database is informed if an anti-allergenics is applied:*

$$
\begin{aligned}
r_6\colon\ (4 : \mathit{give\_antiallergenic}) &\leftarrow (4 : \mathit{need\_strong}), \\
&\qquad (1 : \mathit{allergy\_strong\_ab}), \mathbf{not}\ (1 : \mathit{allergy\_blocked}). \\
r_7\colon\ (1 : \mathit{allergy\_blocked.}) &\leftarrow (4 : \mathit{give\_antiallergenic}).
\end{aligned}
$$

*The resulting system has two minimal inconsistency explanations: the previous explanation* $(\{r_1, r_2, r_4\}, \{r_5\})$*, and the new* $(\{r_1, r_2, r_4, r_6, r_7\}, \{r_6, r_7\})$*. The latter show the typical effect of an odd cycle: both rules of the odd cycle,* $r_6$ *and* $r_7$*, are present in both components of the minimal explanation. Intuitively, all rules of the cycle are necessary to cause the inconsistency while founding the cycle anywhere prevents the inconsistency. Minimal diagnoses of this MCS are* $(\{r_1\}, \emptyset)$*,* $(\{r_2\}, \emptyset)$ $(\{r_4\}, \emptyset)$*,* $(\{r_6\}, \{r_5\})$*,* $(\{r_7\}, \{r_5\})$*,* $(\emptyset, \{r_5, r_6\})$*, and* $(\emptyset, \{r_5, r_7\})$*.*

**Example 11.** *To show that explanations separate independent reasons for inconsistency, and that they report only relevant inconsistencies, consider $M_a = (C_{b1}, C_{a2}, C_{a3}, C_{a4}, C_{a5})$ depicted in Figure 2a. All contexts use logic $L^{asp}_{\Sigma}$ from Example 4 with $\Sigma = \{a, b, \dots, z\}$. This system is inconsistent, because $u$ is a fact in $C_{a4}$ and therefore $r_{a4}$ adds fact $t$ to $C_{a5}$ which makes $C_{a5}$ inconsistent. An alternative source of inconsistency is that $z$ is a fact in $C_{a3}$, therefore $r_{a3}$ adds fact $w$ to $C_{a4}$ which makes $C_{a4}$ inconsistent. $E^{\pm}_m(M)$ contains only the explanations $(\{r_{a3}\}, \emptyset)$ and $(\{r_{a4}\}, \emptyset)$, which each capture one source of inconsistency. Note that that $M_a[\{r_{a2}\}]$ also is an inconsistent system, because $r_{a2}$ adds the fact $w$ to $C_{a4}$, which makes $C_{a4}$ inconsistent. But, since $M_a[\{r_{a1}, r_{a2}\}]$ is consistent, this inconsistency is not relevant and therefore not reported by our notions.*

**Example 12.** *This example shows that mutually exclusive bridge rules can be part of the same explanation, and some advantage of subset- over cardinality-minimality. Consider the MCS $M_b = (C_{b1}, C_{b2}, C_{b3}, C_{b4})$ depicted in Figure 2b. Again, all contexts use logic $L^{asp}_{\Sigma}$ from Example 4 with $\Sigma = \{a, b, \dots, z\}$. $M_b$ is inconsistent, as $p$ causes inconsistency in $C_{b4}$ and $p$ is added to the knowledge base of $C_{b4}$ by bridge rule $r_{b3}$. Due to bridge rule $r_{b1}$, $r$ is always believed by $C_{b2}$, hence $r_{b3}$ is always applicable. This inconsistency cannot be prevented by bridge rules of $M_b$, or unconditional versions thereof. Therefore $(\{r_{b1}, r_{b3}\}, \emptyset)$ is a minimal explanation of $M_b$. Another minimal explanation is $(\{r_{b2}, r_{b3}, r_{b4}\}, \emptyset)$, where the bodies of $r_{b2}$ and $r_{b3}$ are mutually exclusive. However, only together they ensure that $C_{b4}$ is inconsistent, regardless of whether $r_{b1}$ is present and whether fact $r$ is believed at $C_{b2}$.*

The above example also shows that cardinality-minimal explanations cannot identify all sources of inconsistency, since there are two $\subseteq$-minimal explanations, but only one cardinality-minimal one. Additionally, the set of cardinality-minimal explanations does not point out all bridge rules that must be modified to obtain a consistent system.

### 3.3 Deletion-Diagnoses / Deletion-Explanations.

For domains where removal of bridge rules is preferred to unconditional addition of rules, we specialize $D^{\pm}$ to obtain diagnoses of the form $(D_1, \emptyset)$. By Occam's razor, subset-minimal diagnoses are preferred.

**Definition 7.** *Given an MCS $M$, a deletion-diagnosis of $M$ is a set $D \subseteq br(M)$ such that $M[br(M) \setminus D] \not\models \bot$. The set of all deletion-diagnoses (resp., $\subseteq$-minimal deletion-diagnoses) is $D^{-}(M)$ (resp., $D^{-}_m(M)$).*

**Example 13.** *In our example, $D^{-}_m(M) = \{\{r_1\}, \{r_2\}, \{r_4\}\}$.*

Specializing inconsistency explanations to the first component, i.e., disregarding that rules may be added unconditionally, all explanations are of the form $(E_1, br(M))$.

**Definition 8.** *Given an MCS $M$, a deletion-explanation of $M$ is a set $E \subseteq br(M)$ such that each $R$, where $E \subseteq R \subseteq br(M)$, satisfies $M[R] \models \bot$. The set of all such ($\subseteq$-minimal) explanations is denoted by $E^{+}(M)$, and the set of $\subseteq$-minimal ones by $E^{+}_m(M)$.*

**Example 14.** *The only, and thus also minimal, deletion-explanation in our running example is given by $\{r_1, r_2, r_4\}$.*

## 3.4 Refined Notions of Diagnosis and Explanation

### 3.4.1 Refined Diagnoses

One can generalize Definition 4 to refined changes of bridge rules, such that bridge rules necessary to be applicable, become applicable by only removing some body atoms instead of all. Let $br_{ref}(M)$ denote the set of bridge rules of $M$ where some body literals have been removed, i.e., $br_{ref}(M) = \{hd(r) \leftarrow B. \mid B \subseteq body(r)\}$ (where we identify the body of a bridge rule with the set of its literals). A function $fg : br(M) \rightarrow br_{ref}(M)$ is called a *body-reduction function*; it maps bridge rules to rules where some or no body atoms are removed. In the following, we identify $fg : br(M) \rightarrow br_{ref}(M)$ with the corresponding function $fg : 2^{br}(M) \rightarrow 2^{br_{ref}(M)}$ on sets of bridge rules, i.e., for a set $R \subseteq br(M)$ we have $fg(R) = \{fg(r) \mid r \in R\}$.

**Definition 9.** *A* refined diagnosis *is a triple $(D_1, D_2, fg)$ consisting of sets of bridge rules $D_1, D_2 \subseteq br(M)$ and a body-reduction function $fg : br(M) \rightarrow br_{ref}(M)$, such that the resulting MCS is consistent, i.e., $M[br(M) \setminus D_1 \cup fg(D_2)] \not\models \bot$. The set of all refined diagnoses is denoted by $D^{\pm,r}(M)$.*

Again, by Occam's razor, we seek minimal refined diagnoses. To that end, we seek to change a minimal set of bridge rules and within this set, we seek a minimal change of bridge rule bodies. Therefore, more conservation of body atoms is considered more minimal. Formally, let $fg$ and $fg'$ be two body-reduction functions on $br(M)$, then $fg$ is more conservative than $fg'$, written $fg \leq fg'$, iff for every $r \in br(M)$ holds $body(fg(r)) \supseteq body(fg'(r))$. Furthermore, we write $fg < fg'$ iff $fg \leq fg'$ and $fg \neq fg'$.

A refined diagnosis $(D_1, D_2, fg) \in D^{\pm,r}(M)$ is called *minimal*, iff for every $(D'_1, D'_2, fg') \in D^{\pm,r}(M)$ such that $D'_1 \subseteq D_1$ and $D'_2 \subseteq D_2$ it holds that $D_1 = D'_1, D_2 = D'_2$, and $fg' \not< fg$. The set of all minimal refined diagnoses is denoted by $D^{\pm,r}_m(M)$. Observe that the conservation of the body-reduction functions only comes into play if the sets of bridge rules are minimal.

**Example 15.** *Consider a slight modification of our running example where data from the patient history is only imported in the expert system, if the patient is currently under treatment in the hospital. So bridge rule $r_5$ is changed to*

$$r_5 \colon (4 : allow\_strong\_ab) \leftarrow (1 : under\_treatment(p)), \mathbf{not}\ (1 : allergy\_strong\_ab)$$

*and our patient is at the hospital, i.e., $kb_1 = \{allergy\_strong\_ab, under\_treatment(p)\}$.*

*Let $fg(r_5) = (4 : allow\_strong\_ab) \leftarrow (1 : under\_treatment(p))$ and $fg(r) = r$ for all $r \in br(M)$ with $r \neq r_5$. Then, $(\emptyset, \{r_5\}, fg) \in D^{\pm,r}_m(M)$, since $fg(r_5)$ allows the strong antibiotic if the patient merely is under treatment.*

Note that one could also think of refining rules in $D_1$, i.e., ensuring that a rule in $D_1$ is not applicable by adding additional atoms to its body. But as there are no hints to which atoms should be added, such a process would result in a large and arbitrary search space. For example, adding $\mathbf{not}\ allergy\_strong\_ab$ to $r_2$ would result in

$$r'_2 \colon (3 : has\_marker(p)) \leftarrow (2 : blood\_marker), \mathbf{not}\ (1 : allergy\_strong\_ab).,$$

which would make the MCS of our running example consistent. Nevertheless, such a rule does not convey any meaning beyond making the MCS consistent, therefore we disregard such kind of manipulations.

But even in the case of minimal refined diagnoses, there is little information gain: every minimal diagnosis $(D_1, D_2) \in D^{\pm}_m(M)$, together with a witnessing equilibrium $S_w$ of $(D_1, D_2)$, can be refined to a minimal diagnosis $(D_1, D_2, fg)$ using the following *refine* function. Let $\mathcal{S}$ be the set of belief states of the

MCS $M$, then $\mathit{refine}(D_2, S_w) : 2^{br(M)} \times \mathcal{S} \rightarrow (br(M) \rightarrow br_{\mathit{ref}}(M))$ is given by $(D_2, S_w) \mapsto fg$ where $fg$ is the body-reduction function defined as follows:

$$fg(r) = \begin{cases} hd(r) \leftarrow B. & \text{if } r \in D_2, B \subseteq \mathit{body}(r), S_w \mid\!\!\leadsto hd(r) \leftarrow B., \\ & \text{and for no } B \subset B' \subseteq \mathit{body}(r) \text{ holds } S \mid\!\!\leadsto hd(r) \leftarrow B'; \\ r & \text{otherwise.} \end{cases}$$

Observe that a refined diagnosis $(D_1, D_2, fg)$ obtained in such way also admits the equilibrium $S_w$, as all rules of $fg(D_2)$ are applicable in $S_w$ and therefore all head beliefs of $D_2$ are added to the respective contexts, which results in the same knowledge bases as for $\mathit{cf}(D_2)$.

**Proposition 2.** *A triple $(D_1, D_2, fg)$ is a minimal refined diagnosis of $M$ iff there exists a diagnosis $(D_1, D_2) \in D^{\pm}_m(M)$ and a witnessing equilibrium $S_w$, such that $\mathit{refine}(D_2, S_w) = fg$ and no witnessing equilibrium $S'_w$ exists where $\mathit{refine}(D_2, S'_w) = fg'$ and $fg' < fg$.*

**Example 16.** *Consider Example 15 again. The set of minimal diagnoses is the same as for the running example, in particular $(\emptyset, \{r_5\})$ is a minimal diagnosis. The refinement of this diagnosis can be computed using its (only) witnessing equilibrium*

$$\begin{aligned} S_v = (&\{\mathit{allergy\_strong\_ab}, \mathit{under\_treatment}(p)\}, \\ &\{\mathit{blood\_marker}, \mathit{xray\_pneumonia}\}, \\ &\{\mathit{Pneumonia}(p), \mathit{has\_marker}(p), \mathit{AtypPneumonia}(p)\}, \\ &\{\mathit{need\_ab}, \mathit{need\_strong}, \mathit{allow\_strong\_ab}, \mathit{give\_strong}\}), \end{aligned}$$

*where only the negated literal of $r_5$ is deleted, this is sufficient to make the rule applicable under $S_w$, i.e., $fg(r_5) = (4 : \mathit{allow\_strong\_ab}) \leftarrow (1 : \mathit{under\_treatment}(p))$ and $(\emptyset, \{r_5\}, fg) \in D^{\pm,r}_m(M)$.*

### 3.4.2 Refined Explanations

Similar to diagnoses, it is possible to consider refined modifications of rules (rather than $\mathit{cf}(R_2)$) in Definition 6.

**Definition 10.** *A* refined explanation *is a triple $(E_1, E_2, fg)$ consisting of sets of bridge rules $E_1, E_2 \subseteq br(M)$ and a body-reduction function $fg$, such that $M[R_1 \cup fg'(R_2)] \models \bot$ holds, for every $E_1 \subseteq R_1 \subseteq br(M)$, $R_2 \subseteq br(M)$, and every body-reduction function $fg'$ where $r \in E_2$ implies $\mathit{body}(fg(r)) \subseteq \mathit{body}(fg'(r))$.*

Here, we shift the "prevention of inconsistency" expressed by $E_2$ in Definition 6 to the body-reduction $fg$: we do not add unconditional bridge rules, i.e., from $br(M) \setminus E_2$, but rather consider all body-reductions $fg'$ for which it holds that bridge rules in $E_2$ retain all literals indicated by $fg$.

**Example 17.** *Consider the modified MCS of Example 15 again. A refined explanation is $(E_1, E_2, fg)$ where $E_1 = \{r_1, r_2, r_4\}$, $E_2 = \{r_5\}$, and $fg(r_5) = (4 : \mathit{allow\_strong\_ab}) \leftarrow \textbf{not } (1 : \mathit{allergy\_strong\_ab})$.*

The notion of a refined explanation is a generalization of the notion of explanation and there is a 1-to-1 correspondence between them.

**Proposition 3.** *For an inconsistent MCS $M$, it holds that $(E_1, E_2) \in E^{\pm}(M)$ iff there exists a body-reduction function $fg$ such that $(E_1, E_2, fg)$ is a refined explanation.*

In contrast to diagnoses, an explanation does not admit a witnessing equilibrium. Therefore, we cannot infer from an explanation whether the addition of a reduced version of a bridge rule would yield consistency.

However, this can be achieved considering a transformed MCS: Consider $M = (C_1, \ldots, C_n)$, then $M^r = (C_1, \ldots, C_n, C_\alpha)$ is the transformed MCS where $C_\alpha$ is a context whose acceptable belief states contain exactly those formulas added to it via bridge rules, e.g., $C_\alpha$ uses the logic $L^{asp}$ and an empty knowledge base $kb_\alpha = \emptyset$. Furthermore, the bridge rules of $br(M^r)$ are obtained from $br(M)$ in such a way that every bridge rule $r \in br(M)$ of form (1) is split into a core rule ($r^{(0)}$) and a supplementary rule for each body atom ($r^{(1)}, \ldots, r^{(m)}$). The set $tr(r)$ of transformed rules corresponding to $r$ is then given by:

$$
\begin{array}{lrl}
tr(r) = \{ & r^{(0)}: & (k:s) \leftarrow (C_\alpha:p_1), \ldots, (C_\alpha:p_j), (C_\alpha:p_{j+1}), \ldots, (C_\alpha:p_m). \\
& r^{(1)}: & (C_\alpha:p_1) \leftarrow (c_1:p_1). \\
& & \ldots \\
& r^{(j)}: & (C_\alpha:p_1) \leftarrow (c_j:p_j). \\
& r^{(j+1)}: & (C_\alpha:p_1) \leftarrow \mathbf{not}\ (c_{j+1}:p_{j+1}). \\
& & \ldots \\
& r^{(m)}: & (C_\alpha:p_m) \leftarrow \mathbf{not}\ (c_m:p_m). \quad \}
\end{array}
$$

Finally, $M^r$ contains for each bridge rule of $M$ the corresponding transformed rules, i.e., $br(M^r) = \bigcup_{r \in br(M)} tr(r)$. Note that, for readability, this transformation assumes beliefs of different contexts to be disjoint.

For example, a bridge rule $(c_1{:}h) \leftarrow (c_2{:}a), \mathbf{not}\ (c_3{:}b)$ of $M$ is transformed to bridge rules $(c_1{:}h) \leftarrow (c_\alpha{:}a'), (c_\alpha{:}b').$, $(c_\alpha{:}a') \leftarrow (c_2{:}a).$, and $(c_\alpha{:}b') \leftarrow \mathbf{not}\ (c_3{:}b)$ of $M^r$.

An explanation $(E_1, E_2) \in E^{\pm}(M^r)$ then allows to construct a refined explanation $(E_1, E_2^r, fg)$ for $M$ as follows: For every $r \in br(M)$, it holds that $r \in E_2^r$ iff $tr(r) \cap E_2 \neq \emptyset$. Furthermore, let $sup(r) = \{body(r') \mid r' \in tr(r) \wedge r' \neq r^(0)\}$, then $fg$ is a body-reduction function on $br(M)$ such that $fg(r) = hd(r) \leftarrow sup(r)$ if $r \in E_2$ and $fg(r) = (r)$ otherwise.

For example, if the supplementary rules $(c_\alpha{:}a') \leftarrow (c_2{:}a).$, is in $E_2$, then the removal of the corresponding literal, here $(c_2 : a)$, from the original bridge rule in $M$ contributes to avoiding the explained inconsistency in $M$. Removal of all corresponding literals indicated by $E_2$ yields a change of bridge rule bodies to avoid the explained inconsistency completely.

# 4 Properties

In this section we first show that, to some extent, diagnoses can be converted to explanations and vice versa; specifically, minimal diagnoses and minimal explanations point out the same bridge rules, a property we call duality. We then prove a useful non-intersection property of minimal diagnoses, and show how modularity of an MCS (defined in the spirit of splitting sets of logic programs) is reflected in the structure of its diagnoses and explanations.

## 4.1 Converting between Diagnoses and Explanations

While duality expresses that minimal diagnoses and minimal inconsistency explanations point out the same set of bridge rules, in the following we consider the relationships between these notions in more detail. We show that it is possible to characterize explanations in terms of diagnoses, and vice versa minimal diagnoses in terms of minimal explanations.

For the following theorem we generalize the notion of a hitting set [Reiter, 1987] from sets to pairs of sets. Given a collection $\mathcal{C} = \{(A_1, B_1), \ldots, (A_n, B_n)\}$ of pairs of sets $(A_i, B_i)$, $A_i, B_i \subseteq U$ over a set $U$, a *hitting set of* $\mathcal{C}$ is a pair of sets $(X, Y)$, $X, Y \subseteq U$ such that for every pair $(A_i, B_i) \in \mathcal{C}$, (i) $A_i \cap X \neq \emptyset$ or (ii) $B_i \cap Y \neq \emptyset$. A hitting set $(X, Y)$ of $\mathcal{C}$ is *minimal*, if no $(X', Y') \subset (X, Y)$ is a hitting set of $\mathcal{C}$.

We consider hitting sets over pairs of sets of bridge rules, and denote by $HS_M(\mathcal{C})$ (resp., $minHS_M(\mathcal{C})$) the set of all (resp., all minimal) hitting sets of $\mathcal{C}$ over $U = br(M)$. Note that in particular $HS_M(\emptyset) = \{(\emptyset, \emptyset)\}$, and $HS_M(\{(\emptyset, \emptyset)\}) = \emptyset$.

**Theorem 1.** *For every MCS $M$,*

*(a) a pair $(E_1, E_2)$ with $E_1, E_2 \subseteq br(M)$ is an inconsistency explanation of $M$ iff $(E_1, E_2) \in HS_M(D^{\pm}(M))$, i.e., $(E_1, E_2)$ is a hitting set of $D^{\pm}(M)$; and*

*(b) a pair $(E_1, E_2)$ with $E_1, E_2 \subseteq br(M)$ is a minimal inconsistency explanation of $M$ iff $(E_1, E_2) \in minHS_M(D^{\pm}(M))$, i.e., $(E_1, E_2)$ is a minimal hitting set of $D^{\pm}(M)$.*

Clearly, a hitting set of a collection $X$ is the same as a hitting set of the collection of the $\subseteq$-minimal elements in $X$; from Theorem 1. we therefore immediately obtain the following.

**Corollary 1.** *For every MCS $M$,*

*(a) a pair $(E_1, E_2)$ with $E_1, E_2 \subseteq br(M)$ is an inconsistency explanation of $M$ iff $(E_1, E_2) \in HS_M(D^{\pm}_m(M))$; and*

*(b) a pair $(E_1, E_2)$ with $E_1, E_2 \subseteq br(M)$ is a minimal inconsistency explanation of $M$ iff $(E_1, E_2) \in minHS_M(D^{\pm}_m(M))$.*

For our next result, we use the following generalization of a well-known result for minimal hitting sets (see [Berge, 1989]).

**Lemma 1.** *For every collection $X = \{X^1, \ldots, X^n\}$ of pairs $X^i = (X^i_1, X^i_2)$ of sets, $1 \leq i \leq n$, such that $X$ is an anti-chain wrt. $\subseteq$, i.e., elements in $X$ are pairwise incomparable ($X^i \subseteq X^j$ with $1 \leq i, j \leq n$ implies $X^i = X^j$) it holds that $minHS_M(minHS_M(X)) = X$.*

Combined with Corollary 1 (b) we thus obtain.

**Theorem 2.** *A pair $(D_1, D_2)$ with $D_1, D_2 \subseteq br(M)$ is a minimal diagnosis of $M$ iff $(D_1, D_2)$ is a minimal hitting set of $E^{\pm}_m(M)$, formally $D^{\pm}_m(M) = minHS_M(E^{\pm}_m(M))$.*

As for computation, Theorem 1 provides a way to compute the set of explanations $E^{\pm}(M)$ from the set $D^{\pm}(M)$ of diagnoses, while Theorem 2 allows us to compute the set $D^{\pm}_m(M)$ of minimal diagnoses from the set of minimal explanations $E^{\pm}_m(M)$. Corollary 1 shows that, for computing $E^{\pm}(M)$ and $E^{\pm}_m(M)$, it is sufficient to know the set $D^{\pm}_m(M)$ of minimal diagnoses.

Note that Theorem 2 generalizes a result of Reiter's approach to diagnosis [Reiter, 1987], since the former describes relationships between minimal hitting sets in a sense similar to the relationship between diagnoses and conflict sets of the latter.

In contrast, note that Theorem 1 (a) uses hitting sets without the requirement of $\subseteq$-minimality.

**Example 18.** *In our running example, we had $E^{\pm}_m(M) = \{(\{r_1, r_2, r_4\}, \{r_5\})\}$ and $D^{\pm}_m(M) = \{(\{r_1\}, \emptyset), (\{r_2\}, \emptyset), (\{r_4\}, \emptyset), (\emptyset, \{r_5\})\}$. An explanation $(E_1, E_2)$ has a nonempty intersection $E_1 \cap D_1 \neq \emptyset$ or $E_2 \cap D_2 \neq \emptyset$ with every minimal diagnosis $(D_1, D_2)$. We thus obtain exactly one minimal explanation $E = (\{r_1, r_2, r_4\}, \{r_5\})$ by Corollary 1; furthermore, all component-wise supersets of $E$ are explanations, as they also hit every minimal diagnosis, e.g. $(\{r_1, r_2, r_3, r_4, r_5\}, \{r_1, r_2, r_3, r_4, r_5\})$, and $(\{r_1, r_2, r_4\}, \{r_1, r_2, r_3, r_4, r_5\})$.*

*For illustrating Theorem 2, consider the single minimal explanation $(E_1, E_2)$ of $M$ with $E_1 = \{r_1, r_2, r_4\}$ and $E_2 = \{r_5\}$. Then any minimal diagnosis $(D_1, D_2)$ must fulfill $E_1 \cap D_1 \neq \emptyset$ or $E_2 \cap D_2 \neq \emptyset$, and there is no smaller pair $(D_1, D_2)$ with that property. This condition holds for all minimal diagnoses in $D^{\pm}_m(M)$, and as they contain singleton sets only, and all rules in $E^{\pm}_m(M)$ have been 'hit' that way, it is easy to see that the condition cannot be true for any smaller pair $(D_1, D_2) \subset (D_1, D_2)$.*

### 4.1.1 Duality

As it appears, explanations and diagnoses point out bridge rules as causes of inconsistency on a dual basis. Intuitively, bridge rules in $E_1$ of an explanation $(E_1, E_2)$ cause inconsistency, while bridge rules in $D_1$ of a diagnosis $(D_1, D_2)$ remove inconsistency; furthermore, adding unconditional forms of bridge rules from $E_2$ spoils inconsisteny, while not adding unconditional forms of bridge rules from $D_2$ spoils consistency.

Both notions point out rules that are erroneous in the way that those rules contribute to inconsistency. This naturally gives rise to the question whether diagnoses and explanations point out the same rules of an MCS as erroneous, or whether they characterize different aspects.

To formalize this question, we introduce relevancy for inconsistency. Given an MCS $M$, a bridge rule $r \in br(M)$ is *relevant for diagnosis* (*d-relevant*) iff there exists a minimal diagnosis $(D_1, D_2)$ of $M$ with $r \in D_1 \cup D_2$. Analogously, $r$ is *relevant for explanation* (*e-relevant*) iff there exists a minimal explanation with $r \in E_1 \cup E_2$.

**Example 19.** *Recall our running example where $D^{\pm}_m(M) = \{(\{r_1\}, \emptyset), (\{r_2\}, \emptyset), (\{r_4\}, \emptyset), (\emptyset, \{r_5\})\}$ while $E^{\pm}_m(M) = \{(\{r_1, r_2, r_4\}, \{r_5\})\}$.*

*Here the set of d-relevant bridge rules is $\{r_1, r_2, r_4, r_5\}$. The set of e-relevant bridge rules is identical to that; in fact, even identical componentwise, i.e.,*

$$\bigcup\{D_1 \mid (D_1, D_2) \in D^{\pm}(M)\} = \{r_1, r_2, r_4\} = \bigcup\{E_1 \mid (E_1, E_2) \in E^{\pm}(M)\}$$

*and*

$$\bigcup\{D_2 \mid (D_1, D_2) \in D^{\pm}(M)\} = \{r_5\} = \bigcup\{E_2 \mid (E_1, E_2) \in E^{\pm}(M)\}.$$

As the following proposition shows, the component-wise coincidence is not accidental. not only are the d-relevant rules exactly the same that are e-relevant, but this even holds if the components of diagnoses and explanations are treated separately. Formalizing this, for any set $X$ of pairs $(A, B)$ we write $\bigcup X$ for $(\bigcup\{A \mid (A, B) \in X\}, \bigcup\{B \mid (A, B) \in X\})$.

**Proposition 4.** *For every inconsistent MCS $M$, $\bigcup D^{\pm}_m(M) = \bigcup E^{\pm}_m(M)$, i.e., the unions of all minimal diagnoses and all minimal inconsistency explanations coincide.*

Proposition 4 is an immediate consequence of the close structural relationships between diagnoses and explanations, which are shown by Theorems 1 and 2.

This provides evidence for our view that both notions capture exactly those parts of an MCS that are relevant for inconsistency, as duality shows that, in total, two very different perspectives on inconsistency state exactly the same parts of the MCS as erroneous.

In practice this allows one to compute the set of all bridge rules which are relevant for making an MCS consistent (i.e., appear in at least one diagnosis) in two ways: either to compute all minimal explanations, or to compute all minimal diagnoses. Furthermore, the duality result allows to exclude, under Occam's razor, all bridge rules that are not part of any diagnosis (or explanation) from further investigation as they can be skipped savely.

Our running example suggests, that duality also holds for deletion-diagnoses and -explanations, which indeed is true:

**Theorem 3.** *For every inconsistent MCS $M$, $\bigcup D^-_m(M) = \bigcup E^+_m(M)$, i.e., the unions of all minimal deletion-diagnoses and all minimal deletion-inconsistency explanations coincide.*

*Proof.* This is a direct consequence of Theorem 4; set in its proof the second components of diagnoses and explanations to $\emptyset$. ☐

### 4.1.2 Asymmetry

We now investigate why it is possible to obtain the set of explanations from the set of diagnoses, while the other direction only works under $\subseteq$-minimality. The following example illustrates this.

**Example 20.** *Consider the MCS $M = (C_1)$ with the ASP context $C_1 = \{\leftarrow a.\}$, and the bridge rules $br(M) = \{r_1 = (1:a) \leftarrow (1:a)., r_2 = (1:a) \leftarrow not(1:b)\}$. Then $D^\pm(M) = \{(\{r_2\}, \emptyset), (\{r_1, r_2\}, \emptyset)\}$, while $E^\pm_m(M) = \{(\{r_2\}, \emptyset)\}$, because only $r_2$ is relevant (cf. Section 3.2) for inconsistency.*

*$E^\pm(M)$ contains all pointwise supersets of $(\{r_2\}, \emptyset)$, viz. $(\{r_2\}, \emptyset)$, $(\{r_1, r_2\}, \emptyset)$, $(\{r_2\}, \{r_1\})$, $(\{r_2\}, \{r_2\})$, $(\{r_1, r_2\}, \{r_1\})$, $(\{r_1, r_2\}, \{r_2\})$, and $(\{r_1, r_2\}, \{r_1, r_2\})$. Now the set of (non-minimal) hitting sets of the set $E^\pm(M)$ of explanations is the set $E^\pm(M)$ itself, while the set $D^\pm(M)$ of diagnoses only contains two elements.*

The reason behind this asymmetry is that the notion of explanation is an order-increasing concept, i.e., all supersets of an explanation are also explanations, while the notion of diagnosis is not, i.e., a superset of a diagnosis is not necessarily a diagnosis.

This difference is due to the fact that explanations characterize only relevant inconsistencies (as discussed in Section 3.2) and by its definition, all supersets of an explanation are explanations. Therefore the set of minimal explanations characterizes the set of explanations. For the notion of diagnosis this is not the case: a system might contain inconsistent bridge rule configurations which do not appear in explanations because they are irrelevant in the original system. Non-minimal diagnoses provide modifications of the system which might cause and at the same time suppress such an irrelevant inconsistency in order to achieve overall consistency.

In summary, a minimal hitting set of the set of diagnoses characterizes the set of minimal explanations (Corollary 1 (b)) and a minimal hitting set of the set of explanations characterizes the set of minimal diagnoses (Theorem 2). With non-minimality it looks different: the non-minimal hitting sets of $D^\pm(M)$ characterize the set $E^\pm(M)$ of explanations (see Theorem 1 (a)), however the non-minimal hitting sets of $E^\pm(M)$ do not characterize the set $D^\pm(M)$ of diagnoses (see Example 20 for a counterexample).

## 4.2 Non-Overlap in Minimal Diagnoses

We conclude a simple but useful property of minimal diagnoses. Definition 4 reveals that, $(D_1, D_2)$ such that $r \in D_2$ is a diagnosis regardless of whether $r \in D_1$. Therefore,

**Proposition 5.** *Every minimal diagnosis $(D_1, D_2)$ of an MCS $M$, fulfills $D_1 \cap D_2 = \emptyset$, i.e., no rule occurs in both components.*

An analog property does not hold for inconsistency explanations; as shown by Example 10: the minimal explanation $(E_1, E_2)$ with $E_1 = \{r_1, r_2, r_4, r_6, r_7\}$ and $E_2 = \{r_6, r_7\}$ is such that $r_6$ and $r_7$ are present in both $E_1$ and $E_2$.

## 4.3 Modularity of Explanations and Diagnoses

We next give a syntactic criterion which enables the computation of explanations for an MCS $M$ in a divide-and-conquer fashion. In particular, minimal explanations of $M$ are then just combinations of the minimal explanations of the smaller parts. Based on the results about conversion between explanations and diagnoses, these results then carry over to diagnoses as well. This can be exploited to compute minimal explanations and minimal diagnoses for certain classes of MCS more efficiently.

An approach to modularization (in particular for hierarchical and partitionable MCS) is that some part does not impact the rest of the system. To this end, we adapt the notion of *splitting set* as introduced by [Lifschitz and Turner, 1994] in the context of logic programming; a splitting set characterizes a subset of a logic program which is independent of other rules in the program by a syntactic property.

Since an MCS may include contexts with arbitrary logics, a purely syntactical criterion can only be obtained by resorting to beliefs occurring in bridge rules, under the implicit assumption that every output belief of a context depends on every input belief of the context. Hence, we split at the level of contexts, i.e., a splitting set is a set of contexts rather than a set of literals.

**Definition 11.** *A set of contexts $U \subseteq c(M)$ is a* splitting set *of an MCS $M$, if every rule $r \in br(M)$ is such that $h_c(r) \in U$ satisfies $b_c(r) \subseteq U$. More formally, $U$ is a splitting set iff $U \supseteq \bigcup\{b_c(r) \mid r \in br(M), h_c(r) \in U\}$.*

*For such $U$, the set $b_U = \{r \in br(M) \mid h_c(r) \in U\}$ is called the* bottom *relative to $U$.*

**Example 21.** *In our running example, we have $c(M) = \{C_1, \ldots, C_4\}$, with e.g., $h_c(r_1) = h_c(r_2) = C_3$, and $b_c(r_1) = b_c(r_2) = \{C_2\}$. So the set $U_1 = \{C_2, C_3\}$ is a splitting set of $M$; its bottom is $b_{U_1} = \{r_1, r_2\}$.*

*The further splitting sets of $M$ are $U_2 = \{C_1\}$ with $b_{U_2} = \emptyset$, $U_3 = \{C_2\}$ with $b_{U_3} = \emptyset$, and $U_4 = \{C_4, C_3, C_2, C_1\}$ with bottom $b_{U_4} = br_M$.*

Intuitively, if $U$ is a splitting set of $M$, then the consistency (respectively inconsistency) of contexts in $U$ does not depend on the contexts in $c(M) \setminus U$. Thus, if $M[b_U]$ is inconsistent, $M$ stays inconsistent (under the assumption that $M[\emptyset] \not\models \bot$).

For a pair $R = (R_1, R_2)$ of sets of bridge rules compatible with $M$ and a set $U$ of contexts we say that $R$ is *$U$-headed* iff $r \in (R_1 \cup R_2)$ implies $h_c(r) \in U$.

**Proposition 6.** *Suppose $U$ is a splitting set of an MCS $M$. Then,*

*(i) $E \in E^{\pm}(M[b_U])$ iff $E \in E^{\pm}(M)$ and $E$ is $U$-headed, and*

*(ii) $D \in D^{\pm}(M[b_U])$ iff there exists some $D' \in D^{\pm}(M)$ such that $D \subseteq D'$.*

**Corollary 2.** *Every minimal explanation of $M[b_U]$ is a minimal explanation of $M$.*

Note that $M[b_U]$ does not yield all explanations that contain rules from $b_U$, but it yields all explanations that contain only rules from $M[b_U]$.

**Example 22.** *Reconsider our running example MCS $M$ from Example 7, where the laboratory database together with the disease ontology forms a splitting set $U = \{C_2, C_3\}$ with $b_U = \{r_1, r_2\}$. Now $M[b_U]$ is consistent, so $E^{\pm}(M[b_U]) = \emptyset$, but the overall MCS is inconsistent with the minimal explanation $E = (\{r_1, r_2, r_4\}, \{r_5\})$. In line with Proposition 6, $E$ contains rules from $b_U$ but $E$ is not $b_U$-headed.*

In the particular case that two splitting sets form a partitioning of the MCS, then both partitions can be treated without considering the other one. This means that explanations only contain rules from one partition and diagnoses of the whole MCS are obtained by simply combining diagnoses of each of the partitions.

**Proposition 7.** *Suppose that both, $U$ and $U' = c(M) \setminus U$, are splitting sets of an MCS $M$. Then, every $E \in E^{\pm}_m(M)$ is either $U$-headed or $U'$-headed.*

**Corollary 3.** *Suppose $U$ and $U' = c(M) \setminus U$ are splitting sets of an MCS $M$. Then, $E^{\pm}_m(M) = E^{\pm}_m(M[b_U]) \cup E^{\pm}_m(M[b_{U'}])$.*

Thus, using $U, U'$ the MCS $M$ can be partitioned into two parts where minimal explanations can be computed independently. From this and Theorem 2 we can conclude that for a partitionable MCS, the set of all minimal diagnoses can be obtained by combining the minimal diagnoses of each partition.

**Proposition 8.** *Suppose that $U$ and $U' = c(M) \setminus U$ are splitting sets of an MCS $M$. Then,*

$$
\begin{aligned}
D^{\pm}_m(M) = \{(A_1 \cup B_1, A_2 \cup B_2) \mid \\
(A_1, A_2) \in D^{\pm}_m(M[b_U]) \text{ and } (B_1, B_2) \in D^{\pm}_m(M[b_{U'}])\}.
\end{aligned}
$$

We combine Example 11 and Example 12 to create an MCS with two partitions and observe that diagnoses and explanations in one partition are independent of the other partition in line with Propositions 7 and 8.

**Example 23** (continued)**.** *Consider $M_a = (C_{a1}, \ldots, C_{a5})$ and $M_b = (C_{b1}, \ldots, C_{b4})$ from Example 11 and 12 (cf. Fig. 2a and 2b). Then $M = (C_{a1}, \ldots, C_{a5}, C_{b1}, \ldots, C_{b4})$ with bridge rules $br(M) = br(M_a) \cup br(M_b)$ has a partitioning $(U, U')$ where $U = \{C_{a1}, \ldots, C_{a5}\}$ and $U' = \{C_{b1}, \ldots, C_{b4}\}$. Then*

$$
\begin{aligned}
E^{\pm}_m(M) =& \{(\{r_{a3}\}, \emptyset), (\{r_{a4}\}, \emptyset), (\{r_{b1}, r_{b3}\}, \emptyset), (\{r_{b2}, r_{b3}, r_{b4}\}, \emptyset)\} \\
=& E^{\pm}_m(M_a) \cup E^{\pm}_m(M_b)
\end{aligned}
$$

*while*

$$
\begin{aligned}
D^{\pm}_m(M) =& \{(\{r_{b1}, r_{b2}, r_{a3}\}, \emptyset), (\{r_{b1}, r_{b2}, r_{a4}\}, \emptyset), (\{r_{b3}, r_{a3}\}, \emptyset), (\{r_{b3}, r_{a4}\}, \emptyset), \\
& (\{r_{b1}, r_{b4}, r_{a3}\}, \emptyset), (\{r_{b1}, r_{b4}, r_{a4}\}, \emptyset)\} \\
=& \{(A_1 \cup B_1, A_2 \cup B_2) \mid (A_1, A_2) \in D^{\pm}_m(M_U), (B_1, B_2) \in D^{\pm}_m(M_{U'})\}.
\end{aligned}
$$

## 5 Computational Complexity

We next consider the complexity of consistency checking, and of diagnosis and explanation recognition in MCS in a parametric fashion. To this end, we recall the complexity classes that we will use, and show that we can abstract an MCS to beliefs used in bridge rules. We use *context complexity* as a parameter to characterize the overall complexity and we establish for hardness generic results for all complexity classes that are closed under conjunction and projection. Table 1 summarizes our results for complexity classes that are typically encountered in knowledge representation.

| Context complexity $\mathcal{CC}(M)$ | Consistency checking MCSEQ | $(A, B) \overset{?}{\in}$ $D^{\pm}(M)$ MCSD | $D^{\pm}_m(M)$ MCSD$_m$ | $E^{\pm}(M)$ MCSE | $E^{\pm}_m(M)$ MCSE$_m$ |
|---|---|---|---|---|---|
| $\mathbf{P}$ | $\mathbf{NP}$ | $\mathbf{NP}$ | $\mathbf{D^P_1}$ | $\mathbf{coNP}$ | $\mathbf{D^P_1}$ |
| $\mathbf{NP}$ | $\mathbf{NP}$ | $\mathbf{NP}$ | $\mathbf{D^P_1}$ | $\mathbf{coNP}$ | $\mathbf{D^P_1}$ |
| $\mathbf{\Sigma^P_i}, i \geq 1$ | $\mathbf{\Sigma^P_i}$ | $\mathbf{\Sigma^P_i}$ | $\mathbf{D^P_i}$ | $\mathbf{\Pi^P_i}$ | $\mathbf{D^P_i}$ |
| $\mathbf{PSPACE}$ | $\mathbf{PSPACE}$ | | | | |
| $\mathbf{EXPTIME}$ | $\mathbf{EXPTIME}$ | | | | |
| Proposition | 9 | 10 | 11 | 12 | 13 |

Table 1: Complexity of consistency checking and recognizing (minimal) diagnoses and explanations, given $(A, B)$ and an MCS $M$ for complexity classes of typical knowledge-representation formalisms. Membership holds for all cases, completeness holds if at least one context is complete for the respective context complexity.

## 5.1 Complexity Classes

Recall that $\mathbf{P}$, $\mathbf{EXPTIME}$, and $\mathbf{PSPACE}$ are the classes of problems that can be decided using a deterministic Turing machine in polynomial time, exponential time, and polynomial space, respectively. Furthermore $\mathbf{NP}$ (resp., $\mathbf{coNP}$) is the class of problems that can be decided on a nondeterministic Turing machine in polynomial time, where one (resp., all) execution paths accept. Recall the polynomial hierarchy, where $\mathbf{\Sigma^P_0} = \mathbf{\Pi^P_0} = \mathbf{P}$, $\mathbf{\Sigma^P_i}$ is $\mathbf{NP}$ with a $\mathbf{\Sigma^P_{i-1}}$ oracle, and $\mathbf{\Pi^P_i}$ is $\mathbf{coNP}$ with a $\mathbf{\Sigma^P_{i-1}}$ oracle.

Given complexity class $C$, we denote by $\mathbf{D}(C)$ the "difference class" of $C$, i.e., $\mathbf{D}(C) = \{L_1 \times L_2 \mid L_1 \in C,\ L_2 \in \mathbf{co}\text{-}C\}$ denotes the complexity class of decision problems that are the "conjunction" of a problem $L_1$ in $C$ and a problem $L_2$ in $\mathbf{co}$-$C$. For example, $\mathbf{D}(\mathbf{NP}) = \mathbf{D^P_1}$ and $\mathbf{D}(\mathbf{\Sigma^P_i}) = \mathbf{D^P_i}$. Deciding whether a pair $(F_1, F_2)$ of a SAT instance $F_1$ and an independent UNSAT instance $F_2$ is a prototypical problem complete for $\mathbf{D^P_1}$. Note in particular that $\mathbf{D}(\mathbf{PSPACE}) = \mathbf{PSPACE}$ and that $\mathbf{D}(\mathbf{EXPTIME}) = \mathbf{EXPTIME}$.

**Closure under Conjunction and Projection** A complexity class $C$ is *closed under conjunction*, if the following holds: given a problem $L$ in $C$, it holds that the problem $L^n$ where $L^n$ is the $n$-fold Cartesian product of $L$, and $I = (I_1, \ldots, I_n)$ is a 'yes' instance of $L^n$ iff every instance $I_j$, $1 \leq j \leq n$ is a 'yes' instances of $L$, is such that $\bigcup_{n \geq 1} L^n$ is also a problem in $C$.

All classes $\mathbf{P}$, $\mathbf{NP}$, $\mathbf{\Sigma^P_i}$, $\mathbf{\Pi^P_i}$, $D(\mathbf{\Sigma^P_i})$, $\mathbf{PSPACE}$, etc. here are closed under conjunction.

A decision problem $L \subseteq \Sigma^\star \times \Sigma^\star$ is *polynomially balanced*, if some polynomial $p$ exists such that $|I'| \leq p(|I|)$ for all $(I, I') \in L$. Moreover, $L$ is a *polynomial projection* of $L' \subseteq \Sigma^\star \times \Sigma^\star$ if $L = \{I \mid \exists I' : (I, I') \in L'\}$ and $L'$ is polynomially balanced (intuitively, $I'$ is a witness of polynomial size for $I$). Given a complexity class $C$, let $\pi(C)$ contain all problems which are a polynomial projection of a problem $L'$ in $C$. Then a complexity class $C$ is *closed under projection* if $\pi(C) \subseteq C$.

The classes $\mathbf{\Sigma_i^P}$, $\mathbf{NP}$, $\mathbf{EXPTIME}$, $\mathbf{PSPACE}$ are closed under projection, while $\mathbf{coNP}$ and $\mathbf{\Pi_i^P}$ are presumably not.

For further background see [Papadimitriou, 1994].

### 5.2 Output-projected Equilibria

Computing equilibria by guessing and verifying so-called "kernels of context belief sets" has been outlined in [Eiter *et al.*, 2009]. For the purpose of recognizing diagnoses and explanations, it suffices to check for consistency, i.e., for existence of an arbitrary equilibrium in an MCS.

Here we first define *output beliefs*, which are the beliefs used in bodies of bridge rules. Then we show that for checking consistency of an MCS, it is sufficient to consider equilibria *projected to output beliefs*.

**Definition 12.** *Given an MCS* $M = (C_1, \ldots, C_n)$*, the* set of output beliefs of $C_i$*,* $OUT_i = \{p \mid \exists(c\,{:}\,p) \in \mathit{body}(r),\ r \in \mathit{br}(M)\}$*, is the set of beliefs p of* $C_i$ *that occur in the bodies of bridge rules.*

**Example 24** (ctd)**.** *In our running example,* $OUT_1 = \{\mathit{allergy\_strong\_ab}\}$*,* $OUT_2 = \{\mathit{xray\_pneumonia}, \mathit{blood\_marker}\}$*,* $OUT_3 = \{d{:}\mathit{BacterialDisease}, d{:}\mathit{AtypPneumonia}\}$*, and* $OUT_4 = \emptyset$*, as no bridge rule contains a belief at context* $C_4$*.*

Using the notion of output beliefs, we let $S_i^o = S_i \cap OUT_i$ be the projection of $S_i$ to $OUT_i$, and for $S = (S_1, \ldots, S_n)$ we let $S^o = (S_1^o, \ldots, S_n^o)$ be the *output-projected belief state* $S^o$ of $S$.

An output-projected belief state provides sufficient information for evaluating the applicability of bridge rules. We next show how to obtain witnesses for equilibria using this projection.

**Definition 13.** *An output-projected belief state* $S^o = (S_1^o, \ldots, S_n^o)$ *of an MCS* $M$ *is an output-projected equilibrium iff for all* $1 \leq i \leq n$*,*

$$S_i^o \in \{T_i^o \mid T_i \in \mathbf{ACC}_i(kb_i \cup \{h_b(r) \mid r \in \mathit{app}(br_i, S^o)\})\}$$

$S^o$ contains information about all (and only about) output beliefs. As these are the beliefs which determine bridge rule applicability, $\mathit{app}(R, S) = \mathit{app}(R, S^o)$; thus we obtain:

**Lemma 2.** *For each equilibrium* $S$ *of an MCS* $M$*,* $S^o$ *is an output-projected equilibrium. Conversely, for each output-projected equilibrium* $S^o$ *of* $M$*, there exists some equilibrium* $T$ *of* $M$ *such that* $T^o = S^o$*.*

Given an MCS $M$, we denote by $\mathrm{EQ}^o(M)$ the set of output-projected equilibria of $M$.

**Example 25** (continued)**.** *In our running example, the equilibrium*

$$\begin{aligned} S = (&\{\mathit{allergy\_strong\_ab}\}, \{\neg \mathit{blood\_marker}, \mathit{xray\_pneumonia}\},\\ &\{d{:}\mathit{Pneumonia}, d{:}\mathit{BacterialDisease}\}, \{\mathit{need\_ab}, \mathit{give\_weak}\}) \end{aligned}$$

*is witnessed by the output-projected equilibrium*

$$\begin{aligned} S^o = (&\{\mathit{allergy\_strong\_ab}\}, \{\mathit{xray\_pneumonia}\},\\ &\{d{:}\mathit{Pneumonia}, d{:}\mathit{BacterialDisease}\}, \emptyset). \end{aligned}$$

*Here we can observe that, for consistency of the overall system, it is not relevant which belief set is accepted at* $C_i$*, only that some belief set is.*

Therefore each equilibrium is witnessed by a single output-projected equilibrium, and each output-projected equilibrium witnesses at least one equilibrium. For consistency checking (i.e., equilibrium existence) in MCS it is therefore sufficient to consider output-projected equilibria.

### 5.3 Context Complexity

The complexity of consistency checking for an MCS clearly depends on the complexity of its contexts. We next define a notion of *context complexity* by considering the roles which contexts play in the problem of consistency checking.

For all complexity considerations, we represent logics $L_i$ of contexts $C_i$ *implicitly*; they are fixed and we do not consider these (possibly infinite) objects to be part of the input of the decision problems we investigate. Accordingly, the instance size of a given MCS $M$ will be denoted by $|M| = |kb_M| + |br(M)|$ where $|kb_M|$ denotes the size of knowledge bases in $M$ and $|br(M)|$ denotes the size of its set of bridge rules.

Consistency of an MCS $M$ can be decided by a Turing machine with input $M$ which (a) guesses an output-projected belief state $S^o \in OUT_1 \times \cdots \times OUT_n$, (b) evaluates the bridge rules on $S^o$, yielding for each context $C_i$ a set of active bridge rule heads $H_i$ wrt. $S^o$, and (c) checks for each context whether it accepts the guessed $S^o_i$ wrt. $H_i$. We call the complexity of step (c) *context complexity*, formalized as follows.

**Definition 14.** *Given a context $C_i = (kb_i, br_i, L_i)$ and a pair $(H, T_i)$, with $H \subseteq IN_i$ and $T_i \subseteq OUT_i$, the* context complexity *$\mathcal{CC}(C_i)$ of $C_i$ is the computational complexity of deciding whether there exists an $S_i \in \mathbf{ACC}_i(kb_i \cup H)$ such that $S_i \cap OUT_i = T_i$.*

**Example 26.** *Contexts with propositional logic $L^c_\Sigma$ (see Example 1) have $\mathbf{D_1^P}$-complete context complexity; indeed, On the other hand, the restricted logic $L^{pl}_\Sigma$, which is used in our running example for contexts $C_1$ and $C_2$ (see Example 2), is tractable; more precisely, the context complexity is $\mathcal{O}(n)$.*

*A relational database can be captured by knowledge bases and belief sets which are sets of tuples in relations. Acceptability of a belief set computes whether a belief set is the closure of a knowledge base wrt. a fixed set of (possibly recursive) Datalog view definitions. Such a context is complete for $\mathbf{P}$ [Dantsin* et al.*, 2001].*

*A propositional answer set program can be captured by a context where knowledge bases are sets of rules and belief sets are sets of propositions. Acceptability of such a context then checks whether a set of propositions is an answer set of a knowledge base. Such a context is complete for $\mathbf{NP}$ [Dantsin* et al.*, 2001]. Similarly, satisfiability checking of Boolean formulas can be captured by $\mathbf{NP}$ contexts.*

*Default Logic programs and disjunctive logic programs (cf. Example 4) have $\mathbf{\Sigma_2^P}$-complete acceptability checking and thus complexity [Dantsin* et al.*, 2001], [Gottlob, 1992].*

*An agent using one of the widely-known modal logics $K_n$, $T_n$, or $S4_n$ with $n \geq 1$ knowledge operators can be represented as a context. Assuming that such a context has knowledge bases and belief sets consisting of formulas, and that the context accepts the closure $\mathcal{C}_X$ of a set of formulas $X$ in the knowledge base, this context is $\mathbf{PSPACE}$-complete [Halpern and Moses, 1992].*

*For contexts hosting ontological reasoning in the Description Logic $\mathcal{ALC}$ (as in Example 3) we have that acceptability checking corresponds to a set of instance checks. As individual instance checking is $\mathbf{EXPTIME}$-complete [Baader* et al.*, 2003] and $\mathbf{EXPTIME}$ is closed under conjunction, such a context is in $\mathbf{EXPTIME}$. For $|OUT_i| = 1$ we see that such a context is also $\mathbf{EXPTIME}$-hard. Therefore a context using logic $L_\mathcal{A}$ has context complexity $\mathbf{EXPTIME}$.*

Given an MCS $M$, we say $M$ has *upper context complexity* $C$, denoted $\mathcal{CC}(M) \leq C$, if $\mathcal{CC}(C_i) \subseteq C$ for every context $C_i$ of $M$; We say $M$ has *lower context complexity* $C$, denoted $\mathcal{CC}(M) \geq C$, if $C \subseteq \mathcal{CC}(C_i)$ for some context $C_i$ of $M$. We say that $M$ has *context complexity* $C$, denoted $\mathcal{CC}(M) = C$, iff $\mathcal{CC}(M) \leq C$ and $\mathcal{CC}(M) \geq C$. That is, if $\mathcal{CC}(M) = C$ all contexts in $M$ have complexity at most $\mathcal{CC}(M)$, and some context in $M$ has $C$-complete complexity, provided the class $C$ has complete problems.

**Example 27** (continued). *In our running example, for $M = (C_1, C_2, C_3, C_4)$ we have $\mathcal{CC}(C_1) = \mathcal{CC}(C_2) = \mathcal{O}(n)$, $\mathcal{CC}(C_3) = \mathbf{EXPTIME}$, and $\mathcal{CC}(C_4) = \mathbf{\Sigma_2^P}$. As $\mathcal{O}(n) \subseteq \mathbf{\Sigma_2^P} \subseteq \mathbf{EXPTIME}$, we obtain $\mathcal{CC}(M) \leq \mathbf{EXPTIME}$, and as $C_2$ is $\mathbf{EXPTIME}$-complete, we obtain $\mathcal{CC}(M) \geq \mathbf{EXPTIME}$; hence $\mathcal{CC}(M) = \mathbf{EXPTIME}$.*

### 5.4 Overview of Complexity Results

We now give an overview of complexity results, and brief intuition about the proofs that are available in the Appendix.

We study the decision problem for consistency (MCSEQ) and recognition problems for diagnoses (MCSD), minimal diagnoses (MCSD$_m$), explanations (MCSE), and minimal explanations (MCSE$_m$). Note that *existence* of diagnoses and explanations is trivial by our basic assumptions that $M$ is inconsistent and that $M[\emptyset]$ is consistent.

Table 1 summarizes our results for context complexities that are present in typical monotonic and non-monotonic KR formalisms. Corresponding theorems are given in Section 5.6, which are more general than the results shown in Table 1.

For a given context complexity $\mathcal{CC}(M)$ of an MCS $M$, MCSEQ has the same computational complexity as MCSD. If the context complexity is $\mathbf{NP}$ or above, this complexity is equal to context complexity; for context complexity $\mathbf{P}$, it is $\mathbf{NP}$. Intuitively, this is explained as follows. For context complexity $\mathbf{NP}$ and above, guessing a belief state and checking whether it is an equilibrium can be incorporated into context complexity without exceeding checking cost; if the context complexity is $\mathbf{P}$, this complexity is $\mathbf{NP}$.

Recognizing minimal diagnoses MCSD$_m$ is complete for the complexity of MCSD, which captures diagnosis recognition, and an additional complementary problem of refuting MCSD, which captures diagnosis minimality recognition. For context complexity $\mathbf{P}$ we have that MCSD$_m$ is $\mathbf{D^P}$-complete.

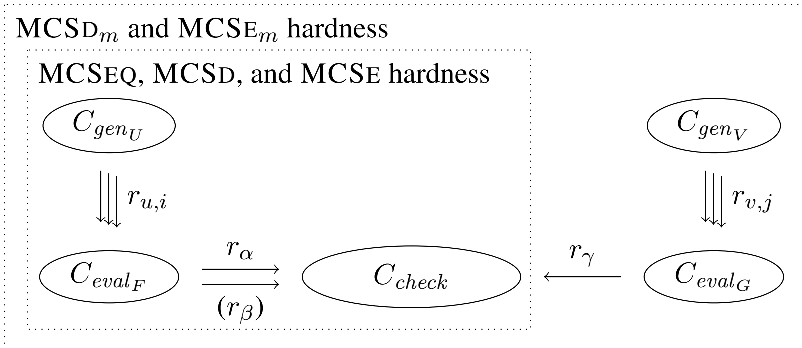
The complexity of MCSE is in the complementary class of the corresponding problem MCSD. Intuitively this is because diagnosis involves existential quantification and explanation involves universal quantification. Accordingly, the complexity of MCSE$_m$ is complementary to MCSD$_m$. As the complexity classes of MCSD$_m$ are closed under complement, MCSE$_m$ and MCSD$_m$ have the same complexity.

These results show that minimal diagnosis and minimal explanation recognition are harder than checking consistency (under usual complexity assumptions), while they are polynomially reducible to each other.

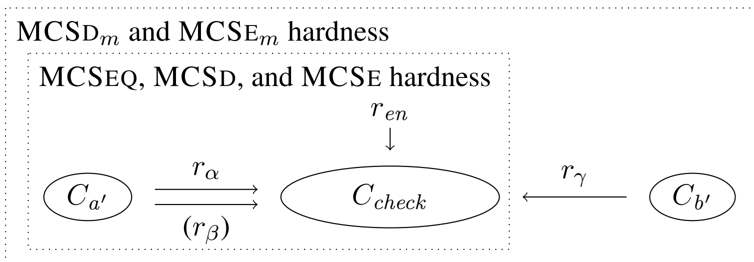
### 5.5 Proof Outline

We treat context complexity of $\mathbf{NP}$ and above uniformly and the case of $\mathbf{P}$ separately. For hardness results we use MCS structures depicted in Figure 3.

For context complexity $\mathbf{P}$ we use reductions from SAT, UNSAT or SAT-UNSAT instances $F$ and/or $G$ to MCS with context complexity $\mathbf{P}$. These reductions use the structure shown in Figure 3a, where contexts $C_{gen_U}$ and $C_{gen_V}$ generate a set of possible truth assignments to sets of variables, $C_{eval_F}$ and $C_{eval_G}$ evaluate formulas $F$ and $G$ under these assignments, and $C_{check}$ checks whether the formulas are satisfiable and/or unsatisfiable. We obtain the hardness via the nondeterministic guess that arises from the different belief sets accepted by contexts $C_{gen_U}$ and $C_{gen_V}$. (See also the description of logic $L_{GUESS}$ in the following.) Our reductions use an acyclic system topology without negation as failure in bridge rules. Note that hardness can also be obtained using a nonmonotonic guess in cyclic bridge rules which contain negation as failure; in that case all contexts of the reduction can be deterministic, i.e., every context accepts at most one belief set for any input. We give such an alternative hardness reduction in the proof of Proposition 9, where we prove $\mathbf{NP}$ hardness of MCSEQ in an MCS of context complexity $\mathbf{P}$.




(a) Structures for lower context complexity $\mathcal{CC}(M) = \mathbf{P}$




(b) Topologies for generic lower context complexity $\mathcal{CC}(M)$

Figure 3: MCS structures for hardness reductions, where dotted areas indicate parts of the MCS used for respective reductions.

Hardness results for context complexity $\mathbf{NP}$ and above are established by a generic reduction: we reduce the problem of acceptability checking of contexts $C_a$ (resp., $C_b$) with context complexity $X$ to decision problems in an MCS $M$ with complexity $X$. These reductions use the scheme shown in Figure 3b, where $C_{a'}$ (resp., $C_{b'}$) evaluates the acceptability checking problem of $C_a$ (resp., $C_b$), and $C_{check}$ tests whether the original problems are "yes" or "no" instances.

For hardness reductions we use the following context logics.

- $L_{ASP}$ is a logic for contexts that contain stratified propositional ASPs with constraints. More in detail, if $L_{ASP} = (\mathbf{BS}, \mathbf{KB}, \mathbf{ACC})$, then $\mathbf{BS}$ is the collection of sets of atoms over a propositional alphabet $\Sigma$, $\mathbf{KB}$ is a set of logic programming rules over $\Sigma$, and given a knowledge base $kb \in \mathbf{KB}$, we define $\mathbf{ACC}(kb) = \mathcal{AS}(kb)$, i.e., the context accepts the set of answer sets of the logic program $kb$. If clear, $\Sigma$ is omitted. In case of stratified propositional ASPs with constraints, a program has at most one answer set. From [Dantsin *et al.*, 2001, Theorem 4.2] it follows that whether an atom $A$ is part of this model is $\mathbf{P}$-complete. Thus, deciding given $OUT_i$ whether $S_i^o \subseteq OUT_i$ is a projected accepted belief set, is $\mathbf{P}$-complete; therefore context complexity is $\mathbf{P}$.

- $L_{GUESS(B)}$ is a trivial logic over the set $B$ that accepts all subsets of its knowledge base. In detail, if logic $L_{GUESS(B)} = (\mathbf{BS}, \mathbf{KB}, \mathbf{ACC})$ then $\mathbf{BS} = \mathbf{KB} = 2^B$ is the powerset of $B$, and $\mathbf{ACC}(kb) = 2^{kb}$ for $kb \in \mathbf{KB}$. If clear, $B$ is omitted. The check whether belief set $S_i^o$ is accepted by knowledge base $kb_i$ can be done in time $\mathcal{O}(|kb_i| + |S_i^o|)$.

### 5.6 Detailed Results

We first formally define the decision problems we consider and then report the complexity results.

**Definition 15.** *Given a MCS $M$,* MCSEQ *is the problem of deciding whether $M$ has an equilibrium.*

**Definition 16.** *Given a MCS $M$ and a pair $(A, B)$ with $A, B \subseteq br(M)$,*

- MCSD *decides whether $(A, B) \in D^{\pm}(M)$, i.e., whether $(A, B)$ is a diagnosis of $M$;*
- MCSD$_m$ *decides whether $(A, B) \in D^{\pm}_m(M)$, i.e., whether $(A, B)$ is a minimal diagnosis of $M$;*
- MCSE *decides whether $(A, B) \in E^{\pm}(M)$, i.e., whether $(A, B)$ is an inconsistency explanation of $M$; and*
- MCSE$_m$ *decides whether $(A, B) \in E^{\pm}_m(M)$, i.e., whether $(A, B)$ is a minimal inconsistency explanation of $M$.*

We next formulate the complexity results.

**Proposition 9.** *The problem* MCSEQ *is*
- $\mathbf{NP}$*-complete if $\mathcal{CC}(M) = \mathbf{P}$, and*
- $C$*-complete if $\mathcal{CC}(M) = C$ and $C$ is a class with complete problems that is closed under conjunction and projection.*

Diagnosis recognition can be done by transforming the MCS using the given diagnosis candidate and deciding MCSEQ. On the other hand, MCSEQ can be reduced to diagnosis recognition of the empty diagnosis candidate $(\emptyset, \emptyset)$. Therefore, diagnosis recognition has the same complexity as consistency checking.

**Proposition 10.** *The problem* MCSD *is*
- $\mathbf{NP}$*-complete if $\mathcal{CC}(M) = \mathbf{P}$, and*
- $C$*-complete if $\mathcal{CC}(M) = C$ and $C$ is a class with complete problems that is closed under conjunction and projection.*

Deciding whether a pair $(A, B)$ is a $\subseteq$-minimal diagnosis of an MCS $M$ requires two checks: (a) whether $(A, B)$ is a diagnosis, and (b) whether no pair $(A', B') \subset (A, B)$ is a diagnosis. The pair $(A, B)$ is a minimal diagnosis iff both checks succeed. This intuitively leads to the following complexity result.

**Proposition 11.** *The problem* MCSD$_m$ *is*
- $\mathbf{D^P_1}$*-complete if $\mathcal{CC}(M) = \mathbf{P}$,*
- $\mathbf{D}(C)$*-complete if $\mathcal{CC}(M) = C$ and $C$ is a class with complete problems that is closed under conjunction and projection.*

Note that, as shown in Table 1, the second item implies that MCSD$_m$ is $\mathbf{D^P_i}$-complete if $\mathcal{CC}(M)$ is complete for $\mathbf{\Sigma^P_i}$ with $i \geq 1$.

Refuting a candidate $(A, B)$ as an explanation of $M$ can be done by guessing a pair of sets $(R_1, R_2)$ from Definition 6 and checking that $M[R_1 \cup cf(R_2)]$ is inconsistent. Then $(A, B)$ is a yes instance iff all guesses succeed, which leads to complementary complexity of consistency checking for that problem. Hardness for context complexity classes $C$ that are closed under conjunction and projection is established via reducing two contexts of complexity $C$ to an MCS which (a) is consistent if both instances are 'yes' instances, (b) has a minimal diagnosis $D$ if both instances are 'no' instances, and (c) has a nonempty minimal diagnosis which is a subset of $D$ if one is a 'yes' and the other a 'no' instance. For context complexity $\mathbf{P}$ a similar approach is used with two SAT instances.

**Proposition 12.** *The problem* MCSE *is*
- $\mathbf{coNP}$*-complete if* $\mathcal{CC}(M) = \mathbf{P}$*, and*
- $\mathbf{co}$-$C$*-complete if* $\mathcal{CC}(M) = C$ *and* $C$ *is a class with complete problems that is closed under conjunction and projection.*

Note that, as shown in Table 1, the second item implies that MCSE is $\mathbf{\Pi_i^P}$-complete if $\mathcal{CC}(M)$ is complete for $\mathbf{\Sigma_i^P}$ with $i \geq 1$.

For complexity results of recognizing minimal explanations we need the following Lemma which limits the number of explanations that need to be checked to verify subset-minimality.

**Lemma 3.** *An explanation* $Q = (Q_1, Q_2)$ *is* $\subseteq$*-minimal iff no pair* $(Q_1, Q_2 \setminus \{r\})$ *with* $r \in Q_2$ *is an explanation and no pair* $(Q_1 \setminus \{r\}, Q_2)$ *with* $r \in Q_1$ *is an explanation.*

Hence, we can check subset-minimality of explanations by deciding whether for linearly many subsets of the candidate $(A, B)$, none is an explanation, i.e., whether for each subset, some $(R_1, R_2)$ exists s.t. $M[R_1 \cup cf(R_2)]$ is consistent. As $\mathbf{NP}$ (resp., $\mathbf{\Sigma_i^P}$) is closed under conjunction and projection, this check is in $\mathbf{NP}$ (resp., $\mathbf{\Sigma_i^P}$). In combination with checking whether the candidate is an explanation, this leads to a complexity of $\mathbf{D_1^P}$ (resp., $\mathbf{D_i^P}$). For context complexity $C \in \mathbf{PSPACE}$ (resp., $C \in \mathbf{EXPTIME}$), $\mathbf{D}(C) = C$. The hardness reduction for MCSE$_m$ is very similar to the one for MCSD$_m$.

**Proposition 13.** *The problem* MCSE$_m$ *is*
- $\mathbf{D_1^P}$*-complete if* $\mathcal{CC}(M) = \mathbf{P}$*,*
- *complete for* $\mathbf{D}(C)$ *if* $\mathcal{CC}(M) = C$ *and* $C$ *is a class with complete problems that is closed under conjunction and projection.*

# 6 Computation

In this section, we show how to compute diagnoses for MCS using HEX-programs and the tool MCS-IE,[1] which is an open source experimental prototype.

First we recall HEX-programs, which extend answer set programs, then show how to compute diagnoses and explanations of MCS, and finally give an overview of the MCS-IE tool.

## 6.1 Preliminaries: HEX-Programs

HEX-programs [Eiter *et al.*, 2005], [Eiter *et al.*, 2006] extend disjunctive logic programs by allowing for access to external information with *external atoms*, and by *predicate variables*.

In this paper, we only use ground (variable-free) HEX-programs and thus recall simplified definitions.

**Syntax** Let $\mathcal{C}$ and $\mathcal{G}$ be mutually disjoint sets of *constants* and *external predicate names*, respectively. Elements from $\mathcal{G}$ are prefixed with "&".

An *ordinary atom* is a formula $p(c_1, \ldots, c_n)$ where $p, c_1, \ldots, c_n$ are constants. An *external atom* is a formula $\&g[\vec{v}](\vec{w})$, where $\vec{v} = Y_1, \ldots, Y_n$ and $\vec{w} = X_1, \ldots, X_m$ are two lists of constants (called *input* and *output* lists, respectively), and $\&g \in \mathcal{G}$ is an external predicate name. Intuitively, an external atom provides a way for deciding the truth value of tuple $\vec{w}$ depending on the extension of input predicates $\vec{v}$.

[1] http://www.kr.tuwien.ac.at/research/systems/mcsie/

A HEX *rule* $r$ is of the form

$$\alpha_1 \vee \ldots \vee \alpha_k \leftarrow \beta_1, \ldots, \beta_m, not\, \beta_{m+1}, \ldots, not\, \beta_n \qquad m, k \geq 0, \tag{2}$$

where all $\alpha_i$ are ordinary atoms and all $\beta_j$ are ordinary or external atoms. Rule $r$ is a *constraint*, if $k = 0$; it is a *fact* if $n = 0$ (in this case we omit $\leftarrow$). A HEX-*program* (or *program*) is a finite set of HEX rules: it is *ordinary*, if it contains only ordinary atoms.

**Semantics** The (ordinary) *Herbrand base* $HB_P^o$ of a HEX-program $P$ is the set of all ordinary atoms $p(c_1, \ldots, c_n)$ occurring in $P$. An *interpretation* $I$ of $P$ is any subset $I \subseteq HB_P^o$; $I$ satisfies (is a *model* of)

- an atom $\alpha$, denoted $I \models \alpha$, if $\alpha \in I$ for an ordinary atom $\alpha$, or if $f_{\&g}(I, \vec{v}, \vec{w}) = 1$ in the case where $\alpha = \&g[\vec{v}](\vec{w})$ and $f_{\&g} : 2^{HB_P^o} \times \mathcal{C}^n \times \mathcal{C}^m \rightarrow \{0, 1\}$ is a (fixed) $(|\vec{v}|+|\vec{w}|+1)$-ary Boolean function associated with $\&g$;
- a rule $r$ of form (2) ($I \models r$), if either $I \models \alpha_i$ for some $\alpha_i$, or $I \models \beta_j$ for some $j \in \{m+1, \ldots, n\}$, or $I \not\models \beta_i$ for some $i \in \{1, \ldots, m\}$;
- a program $P$ ($I \models P$), iff $I \models r$ for all $r \in P$.

The *FLP-reduct* [Faber *et al.*, 2004] of a program $P$ wrt. an interpretation $I$ is the set $fP^I \subseteq P$ of all rules $r$ of form (2) in $P$ such that $I \models \beta_i$, for all $i \in \{1, \ldots, m\}$ and $I \not\models \beta_j$ for all $j \in \{m+1, \ldots, n\}$ (i.e., $I$ satisfies the body of (2)). Then, $I$ is an *answer set of* $P$ iff $I$ is a $\subseteq$-minimal model of $fP^I$. We denote by $\mathcal{AS}(P)$ the collection of all answer sets of $P$.

For $P$ without external atoms, this coincides with answer sets as in [Gelfond and Lifschitz, 1991], for a discussion on the relation between FLP-reduct and GL-reduct see [Faber *et al.*, 2004]. HEX programs can be evaluated using the dlvhex solver. [2] A detailed comparison of HEX programs and MCS, showing similarities and differences, is given in [Eiter *et al.*, 2009].

## 6.2 Computing Diagnoses

We next use HEX programs to describe a generic approach for computing diagnoses, and a way for checking consistency of MCS. In order to compute diagnoses more efficiently, we then integrate both HEX programs.

### 6.2.1 Generic Approach

We can compute diagnoses for some MCS $M$ by guessing a candidate diagnosis and checking whether it yields a consistent system.

We only consider diagnoses $(D_1, D_2)$ where $D_1 \cap D_2 = \emptyset$; diagnoses with $D_1 \cap D_2 \neq \emptyset$ are trivially obtained from these, and they are never minimal (cf. Proposition 5); while we are often interested only in the latter.

Given an MCS $M$, we assemble a HEX-program $P_D(M)$ as follows. For each bridge rule $r \in br(M)$, we add the following guessing rule. Here and in the following, we use $r$ as a name for itself.

$$um(r) \vee d_1(r) \vee d_2(r). \tag{3}$$

[2] http://www.kr.tuwien.ac.at/research/systems/dlvhex/

Intuitively, the predicates $d_1$ and $d_2$ hold bridge rules that are removed from $M$; respectively are added in unconditional form to $M$; $um$ denotes unmodified bridge rules.

Furthermore we create a check for the diagnosis property, which is 'outsourced' to an external atom $\&eq_M[d_1, d_2]()$ with the following evaluation function:

$$f_{\&eq_M}(I, d_1, d_2) = 1 \text{ iff}$$
$$M[br(M) \setminus \{r \mid d_1(r) \in I\} \cup cf(\{r \mid d_2(r) \in I\})] \not\models \bot. \quad (4)$$

Using this external atom, the following constraint eliminates all answer sets that do not correspond to diagnoses:

$$\leftarrow not\ \&eq_M[d_1, d_2](). \quad (5)$$

The program $P_D(M)$ comprising (3) and (5) properly captures diagnoses. The answer sets of $P_D(M)$ correspond to the diagnoses of $M$ as follows.

**Theorem 4.** *Let $M$ be an MCS, then (i) for each answer set $I$ of $P_D(M)$, the pair $(D_{I,1}, D_{I,2}) = (\{r \in br(M) \mid d_1(r) \in I\}, \{r \in br(M) \mid d_2(r) \in I\})$ is a diagnosis of $M$, and (ii) for each diagnosis $(D_1, D_2) \in D^{\pm}(M)$ with $D_1 \cap D_2 = \emptyset$, there exists some answer set $I$ of $P_D(M)$ such that $(D_{I,1}, D_{I,2}) = (D_1, D_2)$.*

Note that to compute all answer sets of this naive encoding, $P_D(M)$, the function $f_{eq_M}$ will be called $3^{|br(M)|}$ times. The encoding we present in the following can drastically reduce the computational effort by doing large parts of the MCS consistency check within the rules part of the HEX encoding. External atoms are used only to evaluate the generic $\mathbf{ACC}$ function of each context in $M$.

### 6.2.2 Consistency Checking

Consistency of an MCS $M$ can be checked by computing output-projected equilibria $S^o$ of $M$. For that, we assemble a program $P_p(M)$ as follows.

We guess presence or absence of each output belief $p$ of each context in $M$.

$$pres_i(p) \vee abs_i(p). \qquad \text{for every } p \in OUT_i,\ 1 \leq i \leq n \quad (6)$$

Given an interpretation $I$ of $P_p(M)$, we use $A_i(I) = \{p \mid pres_i(p) \in I\}$, $1 \leq i \leq n$, to denote the set of output beliefs at context $C_i$, corresponding to the guess in (6).

We evaluate each bridge rule (1) by two corresponding HEX rules, depending on output beliefs guessed in (6).

$$\begin{aligned} in_i(s) \leftarrow & not\ d_1(r), pres_{c_1}(p_1), \ldots, pres_{c_j}(p_j), \\ & not\ pres_{c_{j+1}}(p_{j+1}), \ldots, not\ pres_{c_m}(p_m). \end{aligned} \quad (7)$$

$$in_i(s) \leftarrow d_2(r). \quad (8)$$

Given an interpretation $I$ of $P_p(M)$, we use $B_i(I) = \{s \mid in_i(s) \in I\}$ to denote the set of bridge rule heads at context $C_i$, activated by the output-projected belief state $\mathcal{A}(I) = (A_1(I), \ldots, A_n(I))$. Note that the atoms $d_2(r)$ and $d_1(r)$ will be justified in the integration of $P_p(M)$ and $P_D(M)$. For $P_p(M)$ as stated above, they do not occur in any rule head, therefore (7) will never be deactivated by $d_1(r)$ and (8) will never become applicable.

Finally, we ensure that answer sets of of $P_p(M)$ correspond to output-projected equilibria by checking whether each context $C_i$ accepts the guessed $A_i(I)$ wrt. the set $B_i(I)$ of bridge rule heads activated by bridge rules. For that, we create an external atom $\&con\_out_i[pres_i, b_i]()$ which computes $\mathbf{ACC}_i$ in an external computation. This external atom returns true iff context $C_i$, when given $B_i(I)$, accepts a belief set $S_i$ such that its projection to output-beliefs $OUT_i$ is equal to $A_i(I)$. Formally,

$$f_{\&con\_out_i}(I, pres_i, in_i) = 1 \text{ iff } A_i(I) \in \{S_i^o \mid S_i \in \mathbf{ACC}_i(kb_i \cup B_i(I))\}.$$

We complete $P_p(M)$ by adding the following constraints.

$$\leftarrow not\ \&con\_out_i[pres_i, b_i](). \qquad \text{for every } i \text{ with } 1 \leq i \leq n \tag{9}$$

Let the program $P_p(M)$ comprise the rules (6), (7), (8), and (9). Then the answer sets $I$ of $P_p(M)$ correspond to the output-projected equilibria of $M$ as follows.

**Proposition 14.** *Let $M$ be an MCS, then (i) for each answer set $I$ of $P_p(M)$, the belief state $\mathcal{A}(I)$ is an output-projected equilibrium of $M$, and (ii) for each output-projected equilibrium $S^o$ of $M$ there exists an answer set $I$ of $P_p(M)$ such that $\mathcal{A}(I) = S^o$.*

As the existence of output-projected equilibria characterizes the consistency of MCS (Theorem 2), we obtain the following.

**Corollary 4.** *Given an MCS $M$, $P_p(M)$ has some answer set iff $M$ is consistent.*

#### 6.2.3 Combining diagnosis guess and consistency checking

To implement $f_{\&eq_M}$ in the program $P_D(M)$, we can use $P_p(M)$: given a diagnosis candidate $(D_1, D_2)$, we simply add a representation of it using facts $d_1(X)$ and $d_2(X)$ to $P_p(M)$. The resulting program is equivalent to $P_p(M[br(M) \setminus D_1 \cup cf(D_2)])$; by returning 1 iff it has an answer set, we obtain a faithful implementation of $f_{\&eq_M}$.

However, it is possible (and more efficient) to integrate the programs $P_p(M)$ and $P_D(M)$: let $P_p^D(M)$ be $P_p(M)$ plus all the rules (3). In $P_p^D(M)$ the guess for a diagnosis $d_1$ and $d_2$ directly effects bridge rule evaluation in (7) and (8). The answer sets of $P_p^D(M)$ then correspond to the diagnoses and the output-projected equilibria of the modified/repaired MCS as follows.

**Theorem 5.** *Let $M$ be an MCS, and let $P_p^D(M)$ be as above. Then*

*(i) for each answer set $I$ of $P_p^D(M)$, the pair $(D_{I,1}, D_{I,2}) = (\{r \in br(M) \mid d_1(r) \in I\}, \{r \in br(M) \mid d_2(r) \in I\})$ is a diagnosis of $M$ and $\mathcal{A}(I) = (A_1(I), \ldots, A_n(I))$ is an output-projected equilibrium of $M[br(M) \setminus D_{I,1} \cup cf(D_{I,2})]$; and*

*(ii) for each diagnosis $(D_1, D_2) \in D^{\pm}(M)$ where $D_1 \cap D_2 = \emptyset$, and for each output-projected equilibrium $S^o$ of $M[br(M) \setminus D_1 \cup cf(D_2)]$, there exists an answer set $I$ of $P_p^D(M)$ such that $(D_1, D_2) = (D_{I,1}, D_{I,2})$ and $S^o = \mathcal{A}(I)$.*

Creating one HEX-program that contains both guessing of diagnosis candidates and evaluation of bridge rule semantics requires one level less of HEX indirection compared to the naive approach of using $P_p(M)$ in $f_{\&eq_M}$. This allows to reduce the number of evaluations of the $\mathbf{ACC}$ function in external computations: if one context does not accept its output given its input then the system is globally inconsistent; in that case checking acceptability of other contexts can be omitted, even if different sets of bridge rules are applicable at these contexts. (This omission is not possible in the naive approach.) In Section 6.4 we discuss an implementation of $P_p^D(M)$ and practical results on scalability.

### 6.3 Computing Explanations

We next address computing explanations and present an encoding in HEX. This encoding is more involved since explanations show relevant inconsistencies only and this relevancy requires to check that all pairs of sets of bridge rules in the explanation range yield inconsistent systems. Given an explanation candidate $E = (E_1, E_2) \in 2^{br(M)} \times 2^{br(M)}$, the *explanation range* of $E$ is

$$Rg(E) = \{(R_1, R_2) \mid E_1 \subseteq R_1 \subseteq br(M) \text{ and } R_2 \subseteq br(M) \setminus E_2\}.$$

Intuitively, $Rg(E)$ are "relevant pairs" for $E$. It follows directly from Definition 6 that, $E = (E_1, E_2) \in E^{\pm}(M)$ iff $M[R_1 \cup cf(R_2)] \models \perp$ for all $(R_1, R_2) \in Rg(E)$.

So, the computational complexity of diagnosis recognition is not the same as the one for explanation recognition (for $\mathcal{CC}(M)$ being $\mathbf{P}$ it is $\mathbf{NP}$ versus $\mathbf{coNP}$). In the following we present a direct encoding, $P_P^E(M)$, in HEX using a technique from answer-set programming (cf. [Eiter and Gottlob, 1995] called saturation, [Leone *et al.*, 2006]). We first guess an explanation candidate $E = (E_1, E_2)$ and then ensure via saturation, that for all pairs of sets $(R_1, R_2) \in Rg(E)$ the modified system is inconsistent, i.e., we check for every $(R_1, R_2) \in Rg(E)$ and for every belief state $S$, that some context does not accept $S$ under the bridge rules of $M[R_1 \cup cf(R_2)]$.

For all $r \in br(M)$, $P_P^E(M)$ contains the following rules to guess an explanation candidate.

$$e1(r) \vee ne1(r). \tag{10}$$
$$e2(r) \vee ne2(r). \tag{11}$$

To give some intuition of the saturation technique, assume that $I$ is the (partial) interpretation corresponding to an explanation candidate guessed by the above rules. To check that every $(R_1, R_2) \in Rg(E)$ yields an inconsistent system, saturation is used as follows: via disjunctive rules, $(R_1, R_2) \in Rg(E)$ is guessed as well as a belief state $S$. If $S$ is not an equilibrium for $M[R_1 cf(R_2)]$, then the atom *spoil* is concluded to be true. This in turn leads to the truth of all other atoms that occur in rules to guess $R_1, R_2, S$, and all other atoms that are necessary to check that $S$ is not an equilibrium. The resulting interpretation, $I^\star$, is said to be saturated (or spoiled); formally, it contains $I_{spoil}$, which is given by:

$$\begin{aligned} I_{spoil} = & \{r1(r), nr1(r), r2(r), nr2(r), body(r) \mid r \in br(M)\} \cup \\ & \{in_i(b) \mid r \in br(M) \wedge h_c(r) = i \wedge h_b(r) = b\} \cup \{spoil\} \cup \\ & \bigcup_{a \in OUT_i} \{pres_i(a), abs_i(a)\} \cup \bigcup_{b \in IN_i} \{in_i(b)\}. \end{aligned}$$

Most importantly, $I^\star$ is a maximal model of $fP_P^E(M)^I$ and every other guess for $(R_1, R_2)$ and $S$ will result in the same interpretation $I^\star$, if $S$ is not an equilibrium of $M[R_1 \cup cf(R_2)]$.

On the other hand, if there is a guess for $(R_1, R_2)$ and $S$ such that $S$ is an equilibrium of $M[R_1 \cup cf(R_2)]$, then the corresponding interpretation $I'$ will not be saturated. Since $I^\star$ is a maximal model, it then holds that $I' \subset I^\star$, hence $I^\star$ is not a minimal model of $fP_P^E(M)^I$. Thus, if $I^\star$ is indeed the minimal model of $fP_P^E(M)^I$, then there can not exist such an $I'$, i.e., for all $(R_1, R_2)$ and $S$ it then holds that $S$ is not an equilibrium of $M[R_1 \cup cf(R_2)]$.

Since we are only interested in explanation candidates $E$ where no equilibrium exists for any $(R_1, R_2) \in Rg(E)$, a constraint is added to ensure that only saturated models comprise an answer set, i.e, we ensure that only $I^\star$ may yield an answer set.

To generate $(R_1, R_2) \in Rg(E)$, for every $r \in br(M)$ the following rules are in $P_P^E(M)$:

$$r1(r) :- e1(r). \tag{12}$$

$$r1(r) \vee nr1(r) :- ne1(r). \tag{13}$$

$$nr2(r) :- e2(r). \tag{14}$$

$$r2(r) \vee nr2(r) :- ne2(r). \tag{15}$$

We further guess a belief state of $M$, so $P_P^E(M)$ contains for every $a \in OUT_i$ with $1 \leq i \leq n$ the following rule:

$$pres_i(a) \vee abs_i(a). \tag{16}$$

Recall that $I$ is an answer set of $P_P^E(M)$ iff $I$ is a $\subseteq$-minimal model of $fP_P^E(M)^I$. As we use saturation and external atoms, this can lead to the undesired effect that some $r \in fP_P^E(M)^I$ is unsupported, i.e., for $a$ being the head of $r$ it can happen that $a \in I$ but the body of $r$ is false under $I$ and no other rule's body with head $a$ is true. To avoid this, each bridge rule of $M$ is encoded such that $a \in I$ implies that a correesponding body also evaluates to true. This is achieved by the addition of a unique atom $body(r)$ for each $r \in br(M)$ and further rules ensuring that each literal in the body of $r$ holds if $body(r) \in I$. So, $P_P^E(M)$ contains for each $r \in br(M)$ of form $(i : b) \leftarrow (i_1 : b_1), \ldots, (i_{k-1} : b_{k-1}), not(i_k : b_k), \ldots, not(i_m : b_m)$ the following rules:

$$\begin{aligned} body(r) :- r1(r), pres_{i_1}(b_1), \ldots, pres_{i_{k-1}}(b_{k-1}),& \\ abs_{i_k}(b_k), \ldots, abs_{i_m}(b_m).& \end{aligned} \tag{17}$$

$$r1(r) :- body(r). \tag{18}$$

$$pres_{i_1}(b_1) :- body(r). \tag{19}$$

$$\ldots$$

$$pres_{i_{k-1}}(b_{k-1}) :- body(r). \tag{20}$$

$$abs_{i_k}(b_k) :- body(r). \tag{21}$$

$$\ldots$$

$$abs_{i_m}(b_m) :- body(r). \tag{22}$$

$$in_i(b) :- body(r). \tag{23}$$

$$in_i(b) :- r2(r). \tag{24}$$

Rules (23) and (24) ensure that the head of $r$ is derived if either the body holds, or if $r$ is unconditional, i.e., $r \in R_2$. For the head $(i : b)$ of $r$, let $[(i : b)]$ be the set of bridge rules whose head is the same, i.e., $[(i : b)] = \{r \in br(M) \mid h_c(r) = i \wedge h_b(r) = b\}$. For each head $(i : b)$ of a bridge rule with $[(i : b)] = \{r_1, \ldots, r_k\}$ the following rule of $P_P^E(M)$ ensures that $(i : b)$ is supported:

$$body(r_1) \vee \ldots \vee body(r_k) \vee r2(r_1) \vee \ldots \vee r2(r_k) :- in_i(b). \tag{25}$$

So far $P_P^E(M)$ guesses an explanation candidate $E$, a pair $(R_1, R_2) \in Rg(E)$, a belief state encoded by $pres$ and $abs$, and the beliefs of applicable bridge rule heads are computed. To ensure that $E$ is an explanation it must be the case that for every pair $(R_1, R_2)$ and belief state $S$ some context $C_i$ does not accept $S_i$ given the input encoded by $in_i$. If some context does not accept $S_i$ then a special atom $spoil$ is derived, i.e., if the external atom $\&con\_out'_i[spoil, pres_i, in_i, out_i]()$ is false then $spoil$ is derived. This atom is also derived if the guess of $S$ and $(R_1, R_2)$ is contradictory by itself. So for every $r \in br(M), a \in OUT_i, i \in \{1, \ldots, n\}$ the following rules are in $P_P^E(M)$:

$$spoil :- not\ \&con\_out'_i[spoil, pres_i, in_i](). \tag{26}$$

$$spoil :- r1(r), nr1(r). \tag{27}$$

$$spoil :- r2(r), nr2(r). \tag{28}$$

$$spoil :- pres_i(a), abs_i(a). \tag{29}$$

We slightly extend the external atom $\&con\_out_i[pres_i, in_i]()$ for checking consistency of a context: if $spoil$ is present, then the external atom must be false. This is needed, since a spoiled interpretation $I^\star$ must be a model of the HEX program, which is only guaranteed if the external atom is false in $I^\star$. So, $\&con\_out'_i[spoil, pres_i, in_i]()$ is based on $\&con\_out_i[pres_i, in_i]()$ as follows:

$$f_{\&con\_out'_i}(I, spoil, pres_i, in_i) = 0 \text{ iff } f_{\&con\_out_i}(I, pres_i, b_i) = 0 \vee spoil \in I.$$

To saturate all guesses, we add the following rules, for all $r \in br(M), i \in ci(M), a \in OUT_i, b \in IN_i$, to $P_P^E(M)$:

$$r1(r) :- spoil. \qquad r2(r) :- spoil. \tag{30}$$

$$nr1(r) :- spoil. \qquad nr2(r) :- spoil. \tag{31}$$

$$abs_i(a) :- spoil. \qquad pres_i(a) :- spoil. \tag{32}$$

$$in_i(b) :- spoil. \qquad body(r) :- spoil. \tag{33}$$

As an interpretation $I$ of a program $P$ is only an answer set if it is a minimal model of $fP^I$, it follows that $I$ is not an answer set if there is a model $I'$ of $fP^I$ with $I' \subset I$. If the guess for $(R_1, R_2)$ and the belief state $S$ is not acceptable at context $C_i$, then $spoil$ is derived and saturation takes place, i.e., $I'$ becomes $\subset$-maximal. If, however, some guess for $(R_1, R_2)$ and $S$ yields an equilibrium of $M$, then the corresponding interpretation $I'$ is a subset of the saturated guesses, thus making the explanation candidate no minimal model of its reduct.

To obtain only valid explanations, $P_P^E(M)$ contains the following constraint:

$$:- not\ spoil. \tag{34}$$

It ensures that only saturated interpretations $I^\star$ can be answer sets. But it only is a $\subseteq$-minimal model of $fP_P^E(M)^I$, if no $I' \subset I$ exists, i.e., if all $(R_1, R_2) \in Rg(E)$ yield an inconsistent system. For more details on the saturation technique we refer to [Leone *et al.*, 2000], [Eiter and Polleres, 2003].

The answer sets of $P_P^E(M)$ now exactly encode all explanations of the inconsistent MCS $M$.

**Theorem 6.** *Let $M$ be an inconsistent MCS. Then $(E_1, E_2) \in E^\pm(M)$ iff there exists an answer set $I$ of $P_P^E(M)$ where $E_1 = \{r \mid e1(r) \in I\}$ and $E_2 = \{r \mid e2(r) \in I\}$.*

| | |
|---|---|
| master.hex: | `#context(1,"dlv_asp_context_acc", "kb1.dlv").`<br>`#context(2,"dlv_asp_context_acc", "kb2.dlv").`<br>`#context(3,"ontology_context3_acc", "").`<br>`#context(4,"dlv_asp_context_acc", "kb4.dlv").`<br>`r1: (3:pneum) :- (2:xraypneum).`<br>`r2: (3:marker) :- (2:marker).`<br>`r3: (4:need_ab) :- (3:pneum).`<br>`r4: (4:need_strong) :- (3:atyppneum).`<br>`r5: (4:allow_strong_ab) :- not (1:allergystrong).` |
| kb1.dlv: | `allergystrong.` |
| kb2.dlv: | `marker. xraypneum.` |
| kb4.dlv: | `give_strong v give_weak :- need_ab.`<br>`give_strong :- need_strong.`<br>`give_nothing :- not need_ab, not need_strong.`<br>`:- give_strong, not allow_strong_ab.` |

Figure 4: Examples for MCS topology and knowledge base input files of the MCS-IE tool. These files encode most parts of our running example.

### 6.4 Implementation and Evaluation

We have implemented the rewritings to HEX in the MCS Inconsistency Explainer (MCS-IE) tool[3] described in [Bögl *et al.*, 2010], which is an experimental prototype based on the dlvhex solver. MCS-IE solves the reasoning tasks of enumerating output-projected equilibria, diagnoses, minimal diagnoses, explanations, and minimal explanations of a given MCS.

Contexts can be realized as ASP programs, or by writing a context reasoning module using a C++ interface which allows for implementing arbitrary formalisms that can be captured by MCS contexts.

An online version of MCS-IE is available.[4], which is a useful research tool for quick analysis of inconsistency in small-scale MCS. It requires no installation of additional software on the user side and allows direct editing of bridge rules and context knowledge-bases. A list of showcase MCS allows to directly compute (minimal) diagnoses and (minimal) explanations also for MCS given in this paper.

**Example 28** (ctd)**.** *Figure 4 shows files which encode our running example MCS in the* MCS-IE *input format. Contexts $C_1$, $C_2$, and $C_4$ are formalized in ASP, with knowledge bases* `kb1.dlv`*,* `kb2.dlv`*, and* `kb4.dlv`*, these contexts are evaluated through a* HEX*-plugin for external atoms, which in turn uses the* dlv *solver. On the other hand, ontology reasoning $C_3$ is implemented in C++. For more details about the format and the interface we refer to [Bögl* et al.*, 2010].*

Figure 5 shows the architecture of the MCS-IE system, which is implemented as a plugin to the dlvhex solver. The MCS $M$ at hand is described by the user in a master input file, which specifies all bridge rules and contexts (it may refer to context knowledge base files). Depending on the configuration of MCS-IE, the desired reasoning tasks are solved using one of the three rewritings $P_p(M)$, $P_p^D(M)$, resp. $P^E(M)$, on the input MCS $M$. MCS-IE enumerates answer sets of the rewritten program, and potentially uses a

[3] http://www.kr.tuwien.ac.at/research/systems/mcsie/

[4] http://www.kr.tuwien.ac.at/research/systems/mcsie/tut/

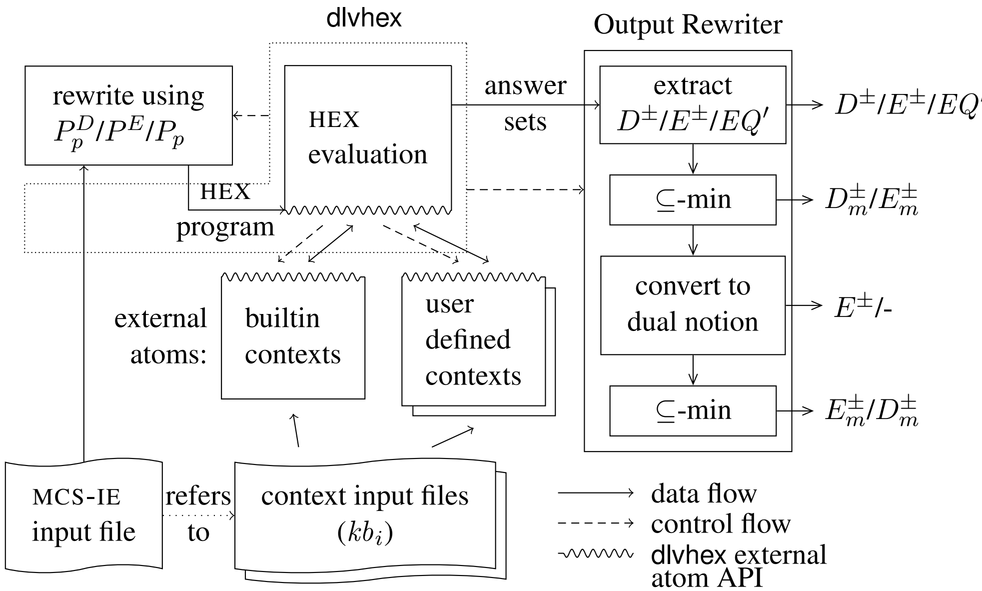



Figure 5: Architecture of the MCS-IE system.

$\subseteq$-minimization module, and a module which realizes the conversions between diagnosis and explanation notions as described in Theorem 2 and Corollary 1. Explanations can be computed by MCS-IE using the direct encoding given in Section 6.3 or through the conversion from diagnoses.

As expected, MCS-IE shows the following behavior wrt. efficiency: the rewriting $P_p^D(M)$, which uses guess-and-check, shows better performance than the rewriting $P^E(M)$, which expresses the $\mathbf{coNP}$ task of recognizing explanations in the $\mathbf{\Sigma_2^P}$ formalism of full-fledged disjunctive HEX programs.

Nevertheless, it appeared that also $P_p^D(M)$ does not scale well. This lead to the development of a better HEX evaluation framework, which divides and conquers the guessing space more efficiently [Eiter *et al.*, 2011a]. While the old evaluation of $P_p^D(M)$ scales exponentially in the total number of output beliefs and bridge rules, the improved one scales exponentially only in the number of output beliefs and bridge rules of the largest context of $M$.

Other approaches to compute diagnoses and explanations of MCS may be faster than the HEX rewriting approach, e.g., distributed evaluation with an extended version of the DMCS algorithm [Bairakdar *et al.*, 2010b]. However, the primary focus of this work are the notions of diagnosis and explanation, investigation of their properties, and an experimental framework for evaluation; therefore the efficient (and more intricate) evaluation methods are left for future work.

# 7 Related Work

Non-monotonicity in MCS was introduced by [Roelofsen and Serafini, 2005] and then further developed ([Brewka *et al.*, 2007], [Brewka and Eiter, 2007]), to eventually allow heterogeneous as well as nonmonotonic systems, and in particular nonmonotonic MCS [Brewka and Eiter, 2007] as considered in this article (cf. [Brewka *et al.*, 2011a] for a more comprehensive account of work related to MCS). However, issues arising from inconsistency of such systems have been largely disregarded.

### 7.0.1 Inconsistency in MCS

A remarkable exception, and thus most closely related to ours, is [Bikakis *et al.*, 2011], where inconsistency in a homogeneous MCS setting is addressed. The approach is to consider defeasible bridge rules for inconsistency removal, i.e., a rule is applicable only if its conclusion does not cause inconsistency. This concept is described in terms of an argumentation semantics in [Bikakis and Antoniou, 2010]. The decision which bridge rules to ignore is based, for every context, on a *strict total order* of all contexts. The set of rules that are ignored thus corresponds to a unique deletion-only diagnosis whose declarative description is more involved compared to our notion, but which is polynomially computable. Note however, that the second component of diagnoses, i.e., rules that are forced to be applicable, have no counterpart in the defeasible MCS inconsistency management approach. Furthermore, the strict total order over contexts forces the user to make (perhaps unwanted) decisions at design time; alternative orders would require a redesign and separate evaluation. Our approach avoids this and can be refined to respect various kinds of orderings and preferences.

Another formalism for homogenous contextualized reasoning that incorporates a form of inconsistency tolerance is the Contextualized Knowledge Repository (CKR) approach [Serafini and Homola, 2012]. It is similar to the MCS approach of formalizing context-dependent knowledge, i.e., a CKR is a set of contexts. Contexts are based on description logic and a hierarchical coverage relation is used to specify that the knowledge of one context, regarding specified topics, is broader than the knowledge of another context. A CKR model then is a collection containing a local DL-model for each context such that constants, concepts and roles that are covered are interpreted exactly the same way in both contexts. The coverage relation itself is specified using a DL-like meta-language.

Consider a DL assertion $P(a)$: if context $C_1$ covers context $C_2$, then the concept $P$ is interpreted in $C_1$ and in $C_2$ in the same way, as well as the individual $a$. MCS are different since the interpretation of $P(a)$ in $C_1$ is not related to that in $C_2$. Similar as in bridge rules of MCS, a CKR context can refer to knowledge from other contexts using a so-called qualifier, e.g., $P(a)_{\{location=Italy,time=2010\}}$ refers to $P(a)$ of a context that covers knowledge about $Italy$ in the year 2010.

A CKR is inconsistency tolerant in the sense that if some context is inconsistent (i.e., its local model is the one with empty domain), then this inconsistency does not propagate to other unrelated contexts. The same property also holds for MCS, but in contrast to CKR, our approach allows to restore consistency by modifying the interlinking of contexts.

Similar in vein to CKR systems are *Modular Ontologies*, i.e., a framework where consistent description logic modules utilize and realize a set of interfaces [Ensan and Du, 2008]. These interfaces are connected by bridge rules for Distributed Description Logic (DDL) [Borgida and Serafini, 2003]. Consistent query answering in a module is achieved by using the maximum consistent set of interfaces utilized by this module only, therefore whole interfaces will be ignored if they would cause any inconsistency in the module. Again, in addition to addressing a more general setting in terms of heterogeneity, our work considers potential modifications of bridge rules that allow to go beyond simple masking of inconsistent parts of the system in order to analyze inconsistency and potentially restore consistency.

Conceptually close to the above homogeneous forms of MCS are *Federated Databases*, a distributed formalism for linked databases [Heimbigner and McLeod, 1985]: objects can be *exported* and *imported* using a decentralized negotiation between two databases. Notably, [Sheth and Larson, 1990] is a survey that, in addition to autonomy (access granting and revoking), is taking up on issues of heterogeneity, however mostly referring to the integration of different query languages. Existing approaches handle incoherence in a database-typical manner of *cascading* or *rejecting* local or distributed constraints. For instance, several protocols for global integrity constraint enforcement are presented in [Grefen and Widom, 1996]. These protocols define the *quiescent* state of the system—when it is *at rest*—and ensure that no constraints are violated in such states. Hence, inconsistency in federated databases is addressed at the level of the (individual) databases rather than their interlinking. Even though resorting to SQL and stratified Datalog allows for non-monotonicity, the possibility of instability in a distributed database system—due to a cyclic dependencies—has not been addressed in the literature. Our work would be suitable to deal with such situations, given that federated databases can be described as MCS with stratified (mostly monotonic) contexts including constraints, and with positive bridge rules.

Concerning the complexity results we established for diagnoses of MCS, we remark that they are related to respective results in abduction: by associating abducible hypotheses with bridge rules, due to the non-monotonicity of the system, recognition of diagnoses corresponds to *cancellation abduction problems*. The latter have been shown to be $\mathbf{NP}$-complete in [Bylander *et al.*, 1991a] under the assumption of a tractable underlying theory (i.e., for $\mathbf{P}$ contexts in our terminology).

For putting our work in a broader context, we subsequently relate it more generally to work on inconsistency management in knowledge bases. We classify and discuss some of the most relevant literature according to the following basic approaches:

- *debugging* techniques serve the purpose of diagnosing information systems, aiming at identifying sources of unexpected and in most cases unintended computation outcomes, and at explaining the latter;

- *repairing* techniques modify the content of knowledge bases in order restore consistency, in particular when new information is incorporated into a knowledge base, or when several knowledge bases are integrated into a single one;

- *consistent query answering* virtually repairs a knowledge base or system, often by ignoring a minimal subset of beliefs or subsystems, and operates on the resulting (virtual) consistent system (i.e., no knowledge is permanently removed);

- *paraconsistent reasoning* accepts contradictory knowledge and, rather than repairing or ignoring (parts of) the information, a more tolerant mode of reasoning is applied that handles also inconsistent pieces of knowledge in a non-trivial way.

Different from most approaches to inconsistency management, our primary aim is to provide a solid theoretical framework for analyzing inconsistency; we do not aim at automatically restoring consistency (although our notions can be used to achieve that).

## 7.1 Debugging in Logic Programming

Debugging in logic programming, i.e., finding out why some logic program has no or an unexpected answer, is remotely related to the problem considered in this paper given that bridge rules look similar to rules of logic programming. A major difference is that in MCS we take contexts with an opaque content into account. In logic programming, presence of an atom in a model of a program directly depends on the firing of rules, which in turn directly depends on the presence or absence of other atoms in the bodies; in the MCS framework, which allows to capture arbitrary logics by abstract belief set functions, there is in general no visible link between the firing of bridge rules and beliefs accepted by a context.

### 7.1.1 Prolog Debugging

A framework for debugging Prolog programs was developed in [Shapiro, 1983]. It relies strongly on the operational specifics of Prolog and consists of a diagnosis and a bug-correction component, where three basic types of errors are considered: (i) termination with incorrect output, (ii) termination with missing output, and (iii) nontermination. For the latter, the approach identifies rules that behave unexpectedly by tracing procedure calls and querying the user whether the procedure call at hand of the form $\langle procedure, input, output\rangle$ is correct. A similar goal is achieved in [Pereira, 1986], where the user should not tell whether such a triple is wrong, but point to a wrong subterm of a procedure call; for that, the implementation builds on a modified unification algorithm that keeps track of the origins of subterms. This is further refined in [Pereira and Calejo, 1988], where the different types of bugs are treated uniformly and by the use of a heuristics the number of questions to the user is reduced.

In comparison, our notion of inconsistency diagnosis roughly corresponds to type (i) and (ii) errors: in a diagnosis $(D_1, D_2)$, $D_1$ contains bridge rules whose head belief is "incorrect", while $D_2$ contains bridge rules whose head belief is "missing". As for (iii), nontermination is not an issue for MCS since no infinite recursion can emerge (modulo computations inside contexts). Furthermore, our approach is fully declarative, without operational attachment adherent to Prolog, and it does not require user input; on the other hand, it only covers consistency and no further aspects. Nonetheless, it is possible to mimic behaviour under user input to some extent by using a technique similar to the meta-reasoning transformation in [Eiter *et al.*, 2010b].

A purely declarative perspective on Prolog debugging is taken in [Lloyd, 1987], based on the formal semantics of extended programs under SLDNF resolution. Again two types of errors are considered, so called "wrong clause instances" (wrong solutions) and "uncovered atoms" (missing solutions). To pinpoint the origin of such errors, the user must specify the intended interpretation of the program, by repeatedly answering queries about the behaviour of the rules.

In [Pereira *et al.*, 1993b] a connection between logic program debugging and abductive diagnosis is investigated. It considers extended logic programs (with strong and default negation) under closed-world assumption (CWA). Based on revisables, i.e., a subset $R$ of the set of literals $not\, L$ assumed true by CWA, and the notion of supported sets $SS(L)$ of a literal $L$, the removal sets of $L$ are defined as the hitting sets of $SS(L)$ restricted to $R$; the ones of the literal $\perp$ indicate how to obtain a non-contradictory program. Using a transformed program $P_1$ of $P$ and information about wrong and missing solutions in $P$, so called minimal revising assumptions (MRAs) of $P_1$ are computed in an iterative manner which identify the reasons for wrong and missing solutions. For programs $P$ that model diagnostic problems, minimal solutions can be obtained from the MRAs.

The ideas and notions in [Pereira *et al.*, 1993b], [Pereira *et al.*, 1993a] are merged in [Lloyd, 1987], [Pereira *et al.*, 1993a] for normal logic programs with constraint rules under well-founded semantics. Referring to them, a diagnosis for a set $U$ of literals is a pair $D = \langle Unc, InR\rangle$ where $Unc$ are uncovered atoms and $InR$ are incorrect rules of $P$, such that $U$ is contained in the well-founded model ($WFM$) of the program $P'$ that results from $P$ by removing all incorrect rules and adding all uncovered atoms. In case of a single minimal diagnosis, the bug in the program is pinpointed precisely; otherwise, the user is asked which diagnosis corresponds to the intended interpretation. This leads to an iterative debugging algorithm that only asks disambiguating queries, i.e., it asks about a subset of the intended interpretation and adds the answer to $U$. Our notion of inconsistency diagnosis, where $D = (D_1, D_2)$ is a diagnosis iff $M[br(M) \setminus D_1 \cup cf(D_2)] \not\models \perp$ resembles this notion for $U = \emptyset$; the underlying semantics of MCS is however very different from $WMF$. Furthermore, there is no counterpart of our inconsistency explanations, nor have refined diagnoses been considered.

#### 7.1.2 ASP Debugging

Answer-set Programming (ASP) is as a rule-based paradigm related to MCS, yet more under grounded equilibrium semantics, which imposes a minimality condition on equilibria [Brewka and Eiter, 2007]; in fact, answer-set programs can be modeled as particular MCS with monotonic rules and non-monotonic birdge rules.

The declarative debugging of answer-set programs was approached by [Syrjänen, 2006] for programs that have no cycles of odd length; in subsequent works, tagging [Brain *et al.*, 2007], meta-programming for ground [Gebser *et al.*, 2008] and non-ground programs [Oetsch *et al.*, 2010], and establishing procedural techniques (breakpoints, step-wise execution) [Oetsch *et al.*, 2011] have been considered. The idea is that an expected answer-set $E$ and an (erroneous) ASP program $P$ are transformed into a program $T$ whose answer-sets explain why $E$ is not an answer-set of $P$. Explanations cover that an instantiation of some rule in $P$ is not satisfied by $E$, as well as the presence of unfounded loops (i.e., lack of foundedness). The latter could be of interest for developing diagnosis of MCS under grounded equilibria semantics; this remains for future work. On the other hand, the procedural techniques seem to be less promising, as MCS lack rule chaining at the abstract level.

A different approach to debug answer-set programs is given in [Balduccini and Gelfond, 2003], where A-Prolog (an ASP-based language) is extended by *consistency-restoring* (CR) rules of the form

$$r: \quad h_1 \, or \ldots or \, h_k \stackrel{+}{\leftarrow} l_1, \ldots, l_m, not \, l_{m+1}, \ldots, not \, l_n.$$

which intuitively reads as: if $l_1, \ldots, l_m$ are accepted beliefs while $l_{m+1}, \ldots, l_n$ are not, then one of $h_1, \ldots, h_k$ "may possibly" be believed. In addition, a preference relation on the rules may be provided. The semantics of CR rules is defined via a translation to abductive logic programs, i.e., logic programs where certain atoms are abducibles (cf. [Kakas *et al.*, 1992]). In answer sets of such programs, a minimal set of abducibles may be assumed to be true without further justification.

Disregarding possible rule preferences, a logic program $P$ with CR rules $CR$ can be embedded to a MCS $M = (C_1)$, where the single context $C_1$ is over disjunctive logic programs, such that the answer sets of $P$ with $CR$ correspond to the witnessing equilibria of the minimal diagnoses $(D_1, D_2)$ of $M$. In more detail, $C_1$ has the knowledge base $kb_1 = P \cup \{cr(r) \mid r \in CR\}$ and bridge rules $br_1 = \{(c_1 : ab(r)) \leftarrow a_\perp. \mid r \in CR\}$, where $a_\perp$ and $ab(r)$ are fresh atoms, for each $r$ as above, and

$$cr(r) = \quad h_1 \vee \ldots \vee h_k \leftarrow ab(r), l_1, \ldots, l_m, not \, l_{m+1}, \ldots, not \, l_n;$$

informally, unconditional firing of a bridge rule simulates the corresponding CR rule; note that $D_1 = \emptyset$.

### 7.2 Content-Based Methods

The methods and approaches underlying research issues and works presented in this subsection exhibit more foundational differences to our approach. Therefore, we will mostly discuss them on a more general level, pointing to some seminal works and survey articles for more extensive coverage of the relevant literature.

#### 7.2.1 Repair Approaches in Integrating Information

A lot of work on inconsistency management has been concentrating on the repair of data when merging, incorporating, or integrating data from different sources. In contrast to our work, in such approaches usually the mappings that relate data of different knowledge bases are fixed, while the contents of the knowledge bases are subject to change in order to restore consistency. This subsumes approaches that do not actually modify original data but modify it virtually (i.e., a view), or operate on a copy.

*Belief revision* and *belief merging* are well understood problems, in particular for classical propositional theories [Konieczny and Pérez, 2011],[Peppas, 2008]. They address how to incorporate a new belief into an existing knowledge base, respectively how to combine knowledge bases, such that the resulting knowledge base is consistent. In this regard, our approach is more related to belief merging than to belief revision. A major difference to belief merging is, however, that MCS connect heterogenous knowledge bases in a decentralized fashion (compared to a centralized merge of uniform knowledge bases), and that selective information exchange among knowledge bases is possible via bridge rules in complex topologies. Furthermore, our work concentrates on changing the mappings between these components in case of conflict, while belief merging strives for modified contents (i.e., knowledge base).

*Abductive reasoning* is often applied to identify pieces of information that need to be changed in order to repair a logical theory or knowledge base, cf. [Inoue and Sakama, 1995], [Lobo and Uzcátegui, 1996], [Zhang and Ding, 2008]. In particular, in [Inoue and Sakama, 1995] abduction is applied to repair theories in (nonmonotonic) logic based on notions of 'explanation' and 'anti-explanation'. Given a theory $K$ and a set $\Gamma$ of abducible formulas, they remove the formulas of a set $O \subseteq \Gamma$, and add the formulas of a set $I \subseteq \Gamma$, to entail (resp. not entail) an observation $F$; i.e., $(K \cup I) \setminus O \models F$ (explanation), resp. $(K \cup I) \setminus O \not\models F$ (anti-explanation). A *repair* of an inconsistent theory $K$ is given by an anti-explanation of $F = \bot$; in particular, if $\Gamma = K$ and $I = \emptyset$, then such a repair is a *maximal consistent subset* of $K$; the use of such sets to restore consistency is central to many approaches of belief revision. Our notion of diagnosis may be regarded as a generalized 2-sorted variant of such anti-explanations, where $O \subseteq \Gamma_O$ and $I \subseteq \Gamma_I$; moreover, under suitable conditions, it is reducible to ordinary anti-explanations. In particular, for $\Gamma_I = \emptyset$ and $\Gamma_O = K$, the maximal consistent subsets of $K$ correspond to the minimal diagnoses of $M_K$. Indeed, for the MCS $M = (C_1)$ with single context $C_1$ having knowledge base $kb_1 = K$ and bridge rules $br_1 = br^{\top}_{\Gamma} \cup br^{\bot}_{\Gamma}$, where $br^{\top}_{\Gamma} = \{(c_1 : \phi) \leftarrow (c_1 : \top) \mid \phi \in \Gamma\}$ and $br^{\bot}_{\Gamma} = \{(c_1 : \phi) \leftarrow (c_1 : \bot) \mid \phi \in \Gamma\}$, then the minimal diagnoses of $M$ correspond to the repairs of $K$. One may replace $\Gamma$ with $\Gamma_O$ in both, $br^{\top}_{\Gamma}$ and $br^{\bot}_{\Gamma}$; furthermore, modified bodies in $br^{\top}_{\Gamma}$ allow for *conditional removal* of formulas, with conditions that might be beyond the expressivness of the language of $K$. As regards explanations, our notion of explanation has no counterpart in the approach of [Inoue and Sakama, 1995].

*Information integration* approaches (see e.g. the papers [Lenzerini, 2002], [Doan and Halevy, 2005], [Leone *et al.*, 2005]) wrap several information sources and materialize the information into one schema. Differences exist in whether the global schema is expressed as a view in terms of the local schemata (global-as-view approaches), or vice versa (local-as-view). The relevant relationships are represented as mappings, which often are specified by database queries; inconsistencies are resolved by modifying the materialized information, thus again by changing contents. However, since the original information sources are not altered, one might consider it closer in spirit to our approach than belief merging. Inconsistency management in information integration systems, and in particular the global-as-view approach, may be regarded as implicit change of mappings, by discarding tuples and/or generating missing tuples. Naturally, this corresponds to deactivating bridge rules and forcing bridge rules to fire, respectively. Different from MCS however, information integration approaches rely on hierarchical, acyclic system topologies and monotone semantics (that can be evaluated using fixpoint algorithms). On the other hand, they apply a more expressive mapping formalism compared to bridge rules in MCS.

*Peer-to-peer data integration* systems [Calvanese *et al.*, 2004] allow for a dynamically changing architecture of a data integration scenario in which peers can enter or leave the system anytime. Inconsistency handling in such systems resorts mainly to approaches that are motivated or akin to techniques of consistent query answering; we thus postpone this to the respective subsection below.

*Ontology mapping* [Choi *et al.*, 2006] and the related tasks of ontology alignment, merging, and integration aim at reusing ontologies in a suitable combination. To this end, mappings between concepts, roles, and individuals are identified to denote the same entity in different ontologies, usually by automatic, statistical methods to 'discover' mappings. They may introduce inconsistency in the (global or local) view on the resulting ontology, even if each individual ontology is consistent. Consistency is achieved by disregarding a mapping if it would add an inconsistency. Heterogeneity in ontology mapping usually refers to different nomenclatures prevailing in different ontologies, or to ontologies in different yet closely related formalisms (e.g., different description logics). In contrast, in MCS heterogeneity refers to combining systems based on different logical formalisms, which need not share any relationship. More notably, however, our work aims at explaining inconsistency, and to provide via diagnoses a more fine-grained possibility to achieve consistency than by simply discarding mappings.

To summarize, the main difference between our work and the contributions to these rather diverse settings of integrating information—and in particular the issue of achieving integrity in doing so—is that we consider modifying the 'mapping', i.e., the interlinking, rather than the data. While the importance of maintaining and repairing mappings has been recognized [Doan and Halevy, 2005], major breakthroughs are still missing.

### 7.2.2 Consistent Query Answering

The approaches considered in this section do not actually modify data to repair an inconsistent system, but virtually consider possible repairs in order to return consistent answers to queries. As this includes (partial) ignorance of information (and thus inconsistency) for the sake of reasoning on a consistent system, the approaches may be regarded as in between repairing and paraconsistent reasoning.

The term *consistent query answering (CQA)* has been coined in the database area where various settings (wrt. integrity constraints and operations for repair) have been considered [Bertossi, 2011], [Bertossi, 2006], [Arenas *et al.*, 2003]. For instance, in the case of denial constraints (including key constraints, functional dependencies, etc.), it is sufficient to restrict the attention to tuple deletions for obtaining repairs and answering queries consistently. Thus, CQA might be regarded as an approach that automatically applies deletion-diagnoses to suppress inconsistent information for answering queries over inconsistent relational databases. Despite this superficial similarity to our work, the differences are apart from heterogeneity much more fundamental: diagnoses and explanations address the interlinking of knowledge bases rather than their content and they aim at making inconsistencies amenable to analysis, explicitly hinting at problems that should be investigated, rather than treating them implicitly for the sake of providing consistent answers. CQA techniques have also been extended to description logic ontologies, e.g., in [Lembo *et al.*, 2010], [Lembo and Ruzzi, 2007], where the taxonomy part is considered to be correct but the data part as possibly inconsistent. Consistent answers to queries are then obtained on maximal consistent subsets of the data wrt. the taxonomy part (and potential further constraints).

Other approaches (but similar in nature) have been applied to answering queries in peer-to-peer data integration settings. An automatic approach for repair was presented in [Calvanese *et al.*, 2008] that ignores inconsistent components, resp. the beliefs held by a minority of peers in the system. Another work on peer-to-peer systems over propositional knowledge bases [Binas and McIlraith, 2008] answers queries over a maximal consistent subset of the knowledge bases. Besides the conceptual difference to MCS regarding the system architecture (dynamic vs. static), our approach explains inconsistency by pointing out mappings that must be changed to achieve consistency. Furthermore, its does not aim at automatic fixes to the system, and in particular not by ignoring entire contexts or beliefs held by a minority among them.

#### 7.2.3 Paraconsistent Approaches

Paraconsistent reasoning approaches (see, e.g., [Hunter, 1996],[Bertossi *et al.*, 2005]) aim upfront at ignoring or tolerating inconsistency in knowledge bases, providing means to reason on them without knowledge explosion, i.e., without justifying arbitrary beliefs (*ex falso quodlibet*); thus, they do not focus on eliminating inconsistency. Nevertheless, in addition to keeping information systems operable in case of inconsistency, paraconsistent reasoning may, similar to our aim, also serve the purpose of analysing inconsistency.

Taking again a very general perspective, in particular disregarding heterogeneity and even the fact that our techniques apply to the interlinking of information, syntactic approaches (e.g. [Besnard and Schaub, 1998]) would be closest to our approach. They essentially restrict theories to the intersection of maximal consistent subsets of formulae as a basis for drawing paraconsistent conclusions. However, while minimal deletion-diagnoses might be viewed as corresponding to maximal consistent subsets, our approach does not prescribe a particular reasoning mode upon them (like considering the system obtained by their intersection). Moreover, our notions of diagnosis and explanation provide more fine-grained structures for analysis than just considering deletion diagnosis, and they deal with nonmonotonic behavior.

The methods that are applied in logic-based approaches to paraconsistent reasoning are completely orthogonal to our techniques. The most prominent representatives resort to *many-valued logics* in order to deal with inconsistency (cf. [Belnap, 1977], [Priest, 1989]). The same applies to *paraconsistent logic programming* [Blair and Subrahmanian, 1989] (see e.g. [Eiter *et al.*, 2010a] for more references and recent works), which therefore also elude themselves from a detailed comparison. Nevertheless, developing model-based techniques for paraconsistent reasoning from inconsistent MCS is an interesting topic for future research. In this regard, [Schenk, 2008] can be inspiring, where trust on information sources on the web has been modeled using an extension of Belnap's four-valued logic [Belnap, 1977] and bridge-rule like constructions based on external predicates govern the information flow.

We conclude this section with a pointer to Gabbay and Hunter [Gabbay and Hunter, 1991] who argued strongly for *managing inconsistency*, in contrast to avoiding, removing, or ignoring it. Their point is that an inconsistent system requires actions to be taken, and in order to do so, different issues need to be taken into account that require a variety of methods. Notably they also developed a corresponding framework in a relational database setting [Gabbay, 1993], [Gabbay and Hunter, 1993]. In this spirit, we consider the notions of diagnosis and inconsistency explanation for MCS as providing a foundational basis for developing methods for more specific tasks on top in order to manage inconsistency of the system.

## 8 Conclusion

We have considered the problem of inconsistency analysis in nonmonotonic Multi-Context Systems (MCS), which are a flexible, abstract formalism to interlink heterogeneous knowledge sources for information exchange. We have presented a consistency-based and an entailment-based notion of inconsistency explanation, called diagnosis and explanation, which are in a duality relation that can be exploited for computational purposes, and which enjoy modularity properties. We have characterized the computational complexity of the two notions, establishing generic results for a range of context complexities. They show that in many cases, explaining inconsistency does not lead to a jump in complexity compared to inconsistency testing, although (unsurprisingly) depending on the interlinking intractability might arise. We have furthermore shown how the notions can be computed by a transformation to HEX programs, which has been implemented in the experimental software tool MCS-IE.

Our results provide a basis for building advanced systems of interlinked knowledge sources, in which the natural need for inconsistency management is supported, by taking specifically the information linkage as a source of inconsistency into account, in contrast to traditional works on inconsistency management that focus on the contents of the knowledge sources; however, in loosely connected systems, control over autonomous knowledge sources is elusive and modifying the information exchange may be the only resort to remove inconsistency.

**Further work** The work presented in this article has been continued in several directions. One of them is to impose different kinds of preferences on the notions of diagnosis and explanation that were introduced here, as in [Eiter *et al.*, 2010b], [Weinzierl, 2012]. They allow for filtering and comparing diagnoses; using meta-programming techniques, the most-preferred ones can be selected from all diagnoses.

Another direction concerns incomplete information about contexts. The setting considered in this article assumes complete information about the behavior of the contexts in information exchange, i.e., for each 'input' of relevant beliefs from other contexts accessed via bridge rules, the 'output' in terms of firing bridge rules is fully known. In real-world applications, however, this information may be only available for specific (classes of) inputs, and querying a context arbitrarily often to gain this knowledge might be infeasible. In such scenarios the notions introduced in [Eiter *et al.*, 2011b] allow to obtain reasonable approximations for diagnoses and explanations of inconsistency.

Finally, another implementation is currently underway in which diagnoses and explanations can be computed by distributed algorithms, exploiting the distributed MCS evaluation framework of [Dao-Tran *et al.*, 2010], [Bairakdar *et al.*, 2010b, Bairakdar *et al.*, 2010a].

**Open Issues** Several issues remain for future work. Building on the notions of preferred diagnosis and explanation, a further topic is to establish concrete inconsistency management procedures for analysis. To this end, a system administrator might ask repeatedly for diagnoses and explanations, considering subsystems and/or a modified interlinkage, and select among the ones presented a most appealing one; the information about past selections may in turn be used to adjust the preferences for calculations.

On the computational side, scalability to scenarios with larger data volume and number of bridge rules is desirable, where the intrinsic complexity of our diagnoses and explanations is prohibitive in general. It remains to single out settings where scalability is still possible, and to get a clearer picture of the scalability frontier. This is linked to the complexity of consistency checking for an MCS; restrictions on the interlinking, in numbers and structure (for the latter, see [Bairakdar *et al.*, 2010b]) will be helpful, as well as properties of the context logics (e.g., monotonicity and unique accepted belief sets). Related to this is developing pragmatic variants of our notions, like focusing by protecting bridge rules (which does not increase worst case complexity), giving up properties (e.g., minimality), or by tolerating inconsistency in parts of the system.

Another issue is to combine inconsistency management of contents and of context interlinking. Recall that Section 7.2 points out how maximal consistent subsets of a knowledge base (which are ubiquitous in content-based inconsistency management) might be simulated using bridge rules. However, an emerging combination—although in a uniform formalism—would be inflexible and less amenable to refinement. More promising is to combine the notions in this article and in [Brewka *et al.*, 2011b], which generalized MCS with a management component for each context and operations to be performed on the knowledge base when a bridge rule fires; this allows for a more sophisticated content-change than simple addition of formulas. Nevertheless, consistency can not be guaranteed in general with such content-based approachs, as inconsistency caused by cyclic information flow can not be resolved. Since the latter can be dealt with by modifying the interlinking, as for instance by our notion of diagnosis, a combination of techniques can be successful.

## References

[Arenas *et al.*, 2003] Marcelo Arenas, Leopoldo E. Bertossi, and Jan Chomicki. Answer sets for consistent query answering in inconsistent databases. *TPLP*, 3(4-5):393–424, 2003.

[Baader *et al.*, 2003] Franz Baader, Diego Calvanese, Deborah McGuinness, Daniele Nardi, and Peter Patel-Schneider, editors. *The Description Logic Handbook: Theory, Implementation and Applications*. Cambridge University Press, Cambridge, UK, 2003.

[Bairakdar *et al.*, 2010a] Seif El-Din Bairakdar, Minh Dao-Tran, Thomas Eiter, Michael Fink, and Thomas Krennwallner. Decomposition of distributed nonmonotonic multi-context systems. In Janhunen and Niemelä [2010], pages 24–37.

[Bairakdar *et al.*, 2010b] Seif El-Din Bairakdar, Minh Dao-Tran, Thomas Eiter, Michael Fink, and Thomas Krennwallner. The DMCS solver for distributed nonmonotonic multi-context systems. In *European Conference on Logics in Artificial Intelligence (JELIA 2010)*, volume 6341, pages 352–355, 2010.

[Balduccini and Gelfond, 2003] Marcello Balduccini and Michael Gelfond. Logic programs with consistency-restoring rules. In *International Symposium on Logical Formalization of Commonsense Reasoning, AAAI 2003 Spring Symposium Series*, pages 9–18, 2003.

[Belnap, 1977] N. D. Belnap. A useful four-valued logic. In G. Epstein and J. M. Dunn, editors, *Modern Uses of Multiple-Valued Logic*, pages 7–37. Reidel Publishing Company, Boston, 1977.

[Berge, 1989] C. Berge. *Hypergraphs*. Elsevier Science Publishers B.V. (North-Holland), Amsterdam, 1989.

[Bertossi *et al.*, 2005] Leopoldo E. Bertossi, Anthony Hunter, and Torsten Schaub, editors. *Inconsistency Tolerance [result from a Dagstuhl seminar]*, volume 3300 of *Lecture Notes in Computer Science*. Springer, 2005.

[Bertossi, 2006] Leopoldo Bertossi. Consistent query answering in databases. *SIGMOD Rec.*, 35(2):68–76, June 2006.

[Bertossi, 2011] Leopoldo E. Bertossi. *Database Repairing and Consistent Query Answering*. Synthesis Lectures on Data Management. Morgan & Claypool Publishers, 2011.

[Besnard and Schaub, 1998] Philippe Besnard and Torsten Schaub. Signed systems for paraconsistent reasoning. *J. Autom. Reasoning*, 20(1):191–213, 1998.

[Bikakis and Antoniou, 2008] Antonis Bikakis and Grigoris Antoniou. Distributed defeasible contextual reasoning in ambient computing. In *Ambient Intelligence*, pages 308–325, 2008.

[Bikakis and Antoniou, 2010] Antonis Bikakis and Grigoris Antoniou. Defeasible contextual reasoning with arguments in ambient intelligence. *IEEE Trans. Knowl. Data Eng.*, 22(11):1492–1506, 2010.

[Bikakis *et al.*, 2011] Antonis Bikakis, Grigoris Antoniou, and Panayiotis Hassapis. Strategies for contextual reasoning with conflicts in ambient intelligence. *Knowl. Inf. Syst.*, 27(1):45–84, 2011.

[Binas and McIlraith, 2008] Arnold Binas and Sheila A. McIlraith. Peer-to-peer query answering with inconsistent knowledge. In *Proceedings of the 11th International Conference on Principles of Knowledge Representation and Reasoning*, pages 329–339, Sydney, Australia, September 16–19 2008.

[Blair and Subrahmanian, 1989] Howard A. Blair and V. S. Subrahmanian. Paraconsistent logic programming. *Theor. Comput. Sci.*, 68(2):135–154, 1989.

[Bögl *et al.*, 2010] Markus Bögl, Thomas Eiter, Michael Fink, and Peter Schüller. The MCS-IE system for explaining inconsistency in multi-context systems. In Tomi Janhunen and Ilkka Niemelä, editors, *12th European Conference on Logics in Artificial Intelligence (JELIA 2010)*, Lecture Notes in Artificial Intelligence, pages 356–359. Springer, September 2010.

[Borgida and Serafini, 2003] A. Borgida and L. Serafini. Distributed description logics: Assimilating information from peer sources. *Journal on data semantics I*, pages 153–184, 2003.

[Brain *et al.*, 2007] Martin Brain, Martin Gebser, Jörg Pührer, Torsten Schaub, Hans Tompits, and Stefan Woltran. Debugging ASP programs by means of ASP. In *Logic Programming and Nonmonotonic Reasoning (LPNMR)*, pages 31–43, 2007.

[Brewka and Eiter, 2007] Gerhard Brewka and Thomas Eiter. Equilibria in heterogeneous nonmonotonic multi-context systems. In *AAAI Conference on Artificial Intelligence (AAAI)*, pages 385–390, 2007.

[Brewka *et al.*, 2007] Gerhard Brewka, Floris Roelofsen, and Luciano Serafini. Contextual default reasoning. In *International Joint Conference on Artificial Intelligence (IJCAI)*, pages 268–273, 2007.

[Brewka *et al.*, 2011a] Gerhard Brewka, Thomas Eiter, and Michael Fink. Nonmonotonic multi-context systems: A flexible approach for integrating heterogeneous knowledge sources. In Marcello Balduccini and Tran Cao Son, editors, *Logic Programming, Knowledge Representation, and Nonmonotonic Reasoning*, volume 6565 of *Lecture Notes in Computer Science*, pages 233–258. Springer, 2011.

[Brewka *et al.*, 2011b] Gerhard Brewka, Thomas Eiter, Michael Fink, and Antonius Weinzierl. Managed multi-context systems. In Toby Walsh, editor, *IJCAI*, pages 786–791. IJCAI/AAAI, 2011.

[Bylander *et al.*, 1991a] T. Bylander, D. Allemang, C. Tanner, and J. Josephson. The computational complexity of abduction. *Artificial Intelligence*, 49:25–60, 1991.

[Bylander *et al.*, 1991b] Tom Bylander, Dean Allemang, Michael C. Tanner, and John R. Josephson. The computational complexity of abduction. *Artif. Intell.*, 49(1-3):25–60, 1991.

[Calvanese *et al.*, 2004] Diego Calvanese, Giuseppe De Giacomo, Maurizio Lenzerini, and Riccardo Rosati. Logical foundations of peer-to-peer data integration. In Catriel Beeri and Alin Deutsch, editors, *PODS*, pages 241–251. ACM, 2004.

[Calvanese *et al.*, 2008] Diego Calvanese, Giuseppe De Giacomo, Domenico Lembo, Maurizio Lenzerini, and Riccardo Rosati. Inconsistency tolerance in P2P data integration: An epistemic logic approach. *Information Systems*, 33(4-5):360–384, 2008.

[Choi *et al.*, 2006] Namyoun Choi, Il-Yeol Song, and Hyoil Han. A survey on ontology mapping. *SIGMOD Rec.*, 35:34–41, September 2006.

[Dantsin *et al.*, 2001] Evgeny Dantsin, Thomas Eiter, Georg Gottlob, and Andrei Voronkov. Complexity and expressive power of logic programming. *ACM Computing Surveys*, 33(3):374–425, 2001.

[Dao-Tran *et al.*, 2010] Minh Dao-Tran, Thomas Eiter, Michael Fink, and Thomas Krennwallner. Distributed nonmonotonic multi-context systems. In Lin et al. [2010], pages 60–70.

[Delgrande and Faber, 2011] James P. Delgrande and Wolfgang Faber, editors. *Logic Programming and Nonmonotonic Reasoning - 11th International Conference, LPNMR 2011, Vancouver, Canada, May 16-19, 2011. Proceedings*, volume 6645 of *Lecture Notes in Computer Science*. Springer, 2011.

[Doan and Halevy, 2005] AnHai Doan and Alon Y. Halevy. Semantic integration research in the database community: A brief survey. *AI Magazine*, 26(1):83–94, 2005.

[Eiter and Gottlob, 1995] Thomas Eiter and Georg Gottlob. On the computational cost of disjunctive logic programming: Propositional case. *Ann. Math. Artif. Intell.*, 15(3-4):289–323, 1995.

[Eiter and Polleres, 2003] Thomas Eiter and Axel Polleres. Transforming co-np checks to answer set computation by meta-interpretation. In Francesco Buccafurri, editor, *APPIA-GULP-PRODE*, pages 410–421, 2003.

[Eiter *et al.*, 2005] Thomas Eiter, Giovambattista Ianni, Roman Schindlauer, and Hans Tompits. A uniform integration of higher-order reasoning and external evaluations in answer-set programming. In *International Joint Conference on Artificial Intelligence*, pages 90–96, 2005.

[Eiter *et al.*, 2006] Thomas Eiter, Giovambattista Ianni, Roman Schindlauer, and Hans Tompits. Effective integration of declarative rules with external evaluations for semantic-web reasoning. In *European Semantic Web Conference (ESWC)*, pages 273–287, 2006.

[Eiter *et al.*, 2009] Thomas Eiter, Gerhard Brewka, Minh Dao-Tran, Michael Fink, Giovambattista Ianni, and Thomas Krennwallner. Combining nonmonotonic knowledge bases with external sources. In *Frontiers of Combining Systems (FroCoS)*, pages 18–42, 2009.

[Eiter *et al.*, 2010a] Thomas Eiter, Michael Fink, and João Moura. Paracoherent Answer Set Programming. In Lin et al. [2010], pages 486–496.

[Eiter *et al.*, 2010b] Thomas Eiter, Michael Fink, and Antonius Weinzierl. Preference-based inconsistency assessment in multi-context systems. In Janhunen and Niemelä [2010], pages 143–155.

[Eiter *et al.*, 2011a] Thomas Eiter, Michael Fink, Giovambattista Ianni, Thomas Krennwallner, and Peter Schüller. Pushing efficient evaluation of HEX programs by modular decomposition. In James Delgrande and Wolfgang Faber, editors, *Logic Programming and Nonmonotonic Reasoning, 11th International Conference (LPNMR 2011)*, pages 93–106, May 2011. (selected for “Best Papers Track” of IJCAI 2011).

[Eiter *et al.*, 2011b] Thomas Eiter, Michael Fink, and Peter Schüller. Approximations for explanations of inconsistency in partially known multi-context systems. In Delgrande and Faber [2011], pages 107–119.

[Ensan and Du, 2008] Faezeh Ensan and Weichang Du. Aspects of inconsistency resolution in modular ontologies. In *Advances in Artificial Intelligence, 21st Conference of the Canadian Society for Computational Studies of Intelligence, Canadian AI 2008, Windsor, Canada, May 28-30, 2008, Proceedings*, volume 5032 of *LNCS*, pages 84–95. Springer, 2008.

[Faber *et al.*, 2004] Wolfgang Faber, Nicola Leone, and Gerald Pfeifer. Recursive aggregates in disjunctive logic programs: Semantics and complexity. In *European Conference on Logics in Artificial Intelligence (JELIA)*, pages 200–212, 2004.

[Gabbay and Hunter, 1991] Dov Gabbay and Anthony Hunter. Making inconsistency respectable 1: A logical framework for inconsistency in reasoning. In *Foundations of Artificial Intelligence Research*, volume 535 of *LNCS*, pages 19–32, 1991.

[Gabbay and Hunter, 1993] Dov M. Gabbay and Anthony Hunter. Making inconsistency respectable: Part 2 - meta-level handling of inconsistency. In *ECSQARU*, volume 747 of *LNCS*, pages 129–136. Springer, 1993.

[Gabbay, 1993] Dov M. Gabbay. Labelled deductive systems: A position paper. In J. Oikkonen and J. Väänänen, editors, *Proc. of Logic Colloquium '90, Helsinki, Finland, 15–22 July 1990*, volume 2 of *Lecture Notes in Logic*, pages 66–88. Springer-Verlag, Berlin, 1993.

[Gebser *et al.*, 2008] Martin Gebser, Jörg Pührer, Torsten Schaub, and Hans Tompits. A meta-programming technique for debugging answer-set programs. In Dieter Fox and Carla P. Gomes, editors, *AAAI*, pages 448–453. AAAI Press, 2008.

[Gelfond and Lifschitz, 1991] Michael Gelfond and Vladimir Lifschitz. Classical negation in logic programs and disjunctive databases. *New Generation Computing*, 9(3/4):365–386, 1991.

[Ghidini and Giunchiglia, 2001] Chiara Ghidini and Fausto Giunchiglia. Local models semantics, or contextual reasoning=locality+compatibility. *Artif. Intell.*, 127(2):221–259, 2001.

[Giunchiglia and Serafini, 1994] Fausto Giunchiglia and Luciano Serafini. Multilanguage hierarchical logics, or: How we can do without modal logics. *Artificial Intelligence*, 65(1):29–70, 1994.

[Giunchiglia, 1993] Fausto Giunchiglia. Contextual reasoning. *Epistemologia*, XVI:345–364, 1993.

[Gottlob, 1992] Georg Gottlob. Complexity results for nonmonotonic logics. *Journal of Logic and Computation*, 2:397–425, 1992.

[Grefen and Widom, 1996] Paul W. P. J. Grefen and Jennifer Widom. Integrity constraint checking in federated databases. In *CoopIS*, pages 38–47, 1996.

[Halpern and Moses, 1992] Joseph Y. Halpern and Yoram Moses. A guide to completeness and complexity for modal logics of knowledge and belief. *Artificial Intelligence*, 54(3):319–379, 1992.

[Heimbigner and McLeod, 1985] Dennis Heimbigner and Dennis McLeod. A federated architecture for information management. *ACM Trans. Inf. Syst.*, 3(3):253–278, 1985.

[Hunter, 1996] Anthony Hunter. *Handbook of Defeasible Reasoning and Uncertain Information*, chapter Paraconsistent Logics, pages 11–36. Kluwer, 1996.

[Inoue and Sakama, 1995] Katsumi Inoue and Chiaki Sakama. Abductive framework for nonmonotonic theory change. In *International Joint Conference on Artificial Intelligence (IJCAI)*, pages 204–210, 1995.

[Janhunen and Niemelä, 2010] Tomi Janhunen and Ilkka Niemelä, editors. *Logics in Artificial Intelligence - 12th European Conference, JELIA 2010, Helsinki, Finland, September 13-15, 2010. Proceedings*, volume 6341 of *Lecture Notes in Computer Science*. Springer, 2010.

[Kakas *et al.*, 1992] Antonis C. Kakas, Robert A. Kowalski, and Francesca Toni. Abductive logic programming. *J. Log. Comput.*, 2(6):719–770, 1992.

[Konieczny and Pérez, 2011] Sébastien Konieczny and Ramón Pino Pérez. Logic based merging. *J. Philosophical Logic*, 40(2):239–270, 2011.

[Lembo and Ruzzi, 2007] Domenico Lembo and Marco Ruzzi. Consistent query answering over description logic ontologies. In *Proc. Conference on Web Reasoning and Rule Systems*, volume 4524 of *LNCS*, pages 194–208. Springer, 2007.

[Lembo *et al.*, 2010] Domenico Lembo, Maurizio Lenzerini, Riccardo Rosati, Marco Ruzzi, and Domenico Fabio Savo. Inconsistency-tolerant semantics for description logics. In Pascal Hitzler and Thomas Lukasiewicz, editors, *RR*, volume 6333 of *Lecture Notes in Computer Science*, pages 103–117. Springer, 2010.

[Lenzerini, 2002] Maurizio Lenzerini. Data integration: A theoretical perspective. In Lucian Popa, Serge Abiteboul, and Phokion G. Kolaitis, editors, *PODS*, pages 233–246. ACM, 2002.

[Leone *et al.*, 2000] Nicola Leone, Riccardo Rosati, and Francesco Scarcello. Enhancing answer set planning. Technical Report DBAI-TR-2000-37, Institut für Informationssysteme – Abteilung Databanken und Artificial Intelligence, Technische Universität Wien, 2000.

[Leone *et al.*, 2005] Nicola Leone, Gianluigi Greco, Giovambattista Ianni, Vincenzino Lio, Giorgio Terracina, Thomas Eiter, Wolfgang Faber, Michael Fink, Georg Gottlob, Riccardo Rosati, Domenico Lembo, Maurizio Lenzerini, Marco Ruzzi, Edyta Kalka, Bartosz Nowicki, and Witold Staniszkis. The INFOMIX system for advanced integration of incomplete and inconsistent data. In *SIGMOD*, pages 915–917, 2005.

[Leone *et al.*, 2006] Nicola Leone, Gerald Pfeifer, Wolfgang Faber, Thomas Eiter, Georg Gottlob, Simona Perri, and Francesco Scarcello. The dlv system for knowledge representation and reasoning. *ACM Trans. Comput. Log.*, 7(3):499–562, 2006.

[Lifschitz and Turner, 1994] Vladimir Lifschitz and Hudson Turner. Splitting a logic program. In *International Conference on Logic Programming (ICLP)*, pages 23–37, 1994.

[Lin *et al.*, 2010] Fangzhen Lin, Ulrike Sattler, and Miroslaw Truszczynski, editors. *Principles of Knowledge Representation and Reasoning: Proceedings of the Twelfth International Conference, KR 2010, Toronto, Ontario, Canada, May 9-13, 2010*. AAAI Press, 2010.

[Lloyd, 1987] J. Lloyd. Declarative error diagnosis. *New Generation Computing*, 5:133–154, 1987. 10.1007/BF03037396.

[Lobo and Uzcátegui, 1996] Jorge Lobo and Carlos Uzcátegui. Abductive change operators. *Fundam. Inform.*, 27(4):385–411, 1996.

[Oetsch *et al.*, 2010] Johannes Oetsch, Jörg Pührer, and Hans Tompits. Catching the ouroboros: On debugging non-ground answer-set programs. *TPLP*, 10(4-6):513–529, 2010.

[Oetsch *et al.*, 2011] Johannes Oetsch, Jörg Pührer, and Hans Tompits. Stepping through an answer-set program. In Delgrande and Faber [2011], pages 134–147.

[Papadimitriou, 1994] Christos H. Papadimitriou. *Computational Complexity*. Addison-Wesley, 1994.

[Peppas, 2008] Pavlos Peppas. *Handbook of Knowledge Representation*, volume 3 of *Foundations of Artificial Intelligence*, chapter Belief Revision, pages 317 – 359. Elsevier, 2008.

[Pereira and Calejo, 1988] Luís Moniz Pereira and Miguel Calejo. A framework for prolog debugging. In *ICLP/SLP*, pages 481–495, 1988.

[Pereira *et al.*, 1993a] Luís Moniz Pereira, Carlos Viegas Damásio, and José Júlio Alferes. Debugging by diagnosing assumptions. In Peter Fritzson, editor, *AADEBUG*, volume 749 of *Lecture Notes in Computer Science*, pages 58–74. Springer, 1993.

[Pereira *et al.*, 1993b] Luís Moniz Pereira, Carlos Viegas Damásio, and José Júlio Alferes. Diagnosis and debugging as contradiction removal. In *LPNMR*, pages 316–330, 1993.

[Pereira, 1986] Lus Moniz Pereira. Rational logic programming, 1986.

[Priest, 1989] Graham Priest. Reasoning about truth. *Artificial Intelligence*, 39(2):231 – 244, 1989.

[Przymusinski, 1991] T.C. Przymusinski. Stable semantics for disjunctive programs. *New generation computing*, 9(3):401–424, 1991.

[Reiter, 1987] R. Reiter. A theory of diagnosis from first principles. *Artificial Intelligence*, 32:57–95, 1987.

[Roelofsen and Serafini, 2005] Floris Roelofsen and Luciano Serafini. Minimal and absent information in contexts. In *International Joint Conference on Artificial Intelligence (IJCAI)*, pages 558–563, 2005.

[Schenk, 2008] Simon Schenk. On the semantics of trust and caching in the semantic web. In Amit P.Sheth et al., editor, *International Semantic Web Conference*, volume 5318 of *Lecture Notes in Computer Science*, pages 533–549. Springer, 2008.

[Serafini and Homola, 2012] Luciano Serafini and Martin Homola. Contextualized knowledge repositories for the semantic web. *J. Web Sem.*, 12:64–87, 2012.

[Shapiro, 1983] Ehud Y. Shapiro. *Algorithmic Program Debugging*. MIT Press, 1983.

[Sheth and Larson, 1990] Amit P. Sheth and James A. Larson. Federated database systems for managing distributed, heterogeneous, and autonomous databases. *ACM Comput. Surv.*, 22(3):183–236, 1990.

[Syrjänen, 2006] Tommi Syrjänen. Debugging inconsistent answer set programs. In *International Workshop on Nonmonotonic Reasoning (NMR)*, pages 77–83, 2006.

[Weinzierl, 2012] Antonius Weinzierl. Comparing inconsistency resolutions in multi-context systems. In Daniel Lassiter and Marija Slavkovik, editors, *New Directions in Logic, Language and Computation*, volume 7415 of *Lecture Notes in Computer Science*, pages 158–174. Springer Berlin Heidelberg, 2012.

[Zhang and Ding, 2008] Yan Zhang and Yulin Ding. Ctl model update for system modifications. *J. Artif. Intell. Res. (JAIR)*, 31:113–155, 2008.

# A Examples

In this section, we give the abstract logics $L_{\mathcal{A}}$ and $L_{\Sigma}^{asp}$ in detail.

**Example 29.** *We formally introduce the abstract logic $L_{\mathcal{A}}$ to capture ontologies and description logic. Over a signature of atomic concepts $\mathbb{C}$, roles $\mathbb{R}$, and individuals $I$, T-Box axioms and A-Box axioms are defined based on the notion of concepts. Concepts are inductively defined as follows: every atomic concept is a concept, and if $C, D$ are concepts and $R \in \mathbb{R}$ is a role, then $C \sqcap D$, $C \sqcup D$, $\neg C$, $\forall R.C$, and $\exists R.C$ are concepts. Given concepts $C, D$, a role $R \in \mathbb{R}$, and individuals $a, b \in I$, a T-Box axiom (terminological axiom) is a formula of the form $C \sqsubseteq D$, and an A-Box axiom (assertional axiom) is either of the form $a : C$, or of the form $(a, b) : R$. Finally, $\mathcal{ALC}$ axioms are either T-Box axioms or A-Box axioms.*

*Then, $L_{\mathcal{A}}$ is composed of*

- $\mathbf{KB}$*, being the collection of sets of $\mathcal{ALC}$ axioms,*
- $\mathbf{BS}$*, being the set of possibly believed assertions, i.e., $\mathbf{BS}$ is the powerset of the set of positive A-Box axioms, and*
- $\mathbf{ACC}$*, being a mapping from knowledge bases to the set of assertions entailed by the knowledge base. As $\mathcal{ALC}$ amounts to a fragment of first-order logic, the semantics of an $\mathcal{ALC}$ knowledge base $\pi$ can be given by a rewriting to first-order logic. For our purpose, $\mathbf{ACC}(\pi) = \{S\}$ where $S$ is the set of classically entailed atomic assertions of the first-order rewriting of $\pi$.*

**Example 30.** *We give the formal definition of $L_{\Sigma}^{asp}$, the abstract logic for disjunctive logic programs under the answer-set semantics over a non-ground signature $\Sigma$. For $L_{\Sigma}^{asp} = (\mathbf{KB}, \mathbf{BS}, \mathbf{ACC})$,*

- $\mathbf{KB}$ *is the set of normal disjunctive logic programs over $\Sigma$, i.e., each $kb \in \mathbf{KB}$ is a set of rules of the form*

  $$a_1 \vee \ldots \vee a_n \leftarrow b_1, \ldots, b_i, \mathit{not}\, b_{i+1}, \ldots, \mathit{not}\, b_m,$$

  *where all $a_i$, $b_j$, are atoms over a first-order language $\Sigma$, and $n + m > 0$. Let $r$ be a rule of the aforementioned form, then $H(r) = \{a_1, \ldots, a_n\}$, $B^+(r) = \{b_1, \ldots, b_i\}$, $B^- = \{b_{i+1}, \ldots, b_m\}$, and $B(r) = B^+(r) \cup B^-(r)$. Each rule $r \in kb$ must be safe, i.e., $vars(H(r)) \cup vars(B^-(r)) \subseteq vars(B^+(r))$, where for a set of atoms $A$, $vars(A) = \{vars(a) \mid a \in A\}$ and $vars(a)$ is the set of first-order variables occurring in the atom $a$,*
- $\mathbf{BS}$ *is the set of Herbrand interpretations over $\Sigma$, i.e, each $bs \in \mathbf{BS}$ is a set of ground atoms from $\Sigma$, and*
- $\mathbf{ACC}(kb)$ *returns the set of $kb$'s answer sets: for $P \in \mathbf{KB}$ and $T \in \mathbf{BS}$ let $P^T = \{r \in grnd(P) \mid T \models B(r)\}$ be the FLP-reduct of $P$ wrt. $T$, where $grnd(P)$ returns the ground version of all rules in $P$. Then $bs \in \mathbf{BS}$ is an answer set, i.e., $bs \in \mathbf{ACC}(kb)$, iff $bs$ is a minimal model of $kb^{bs}$.*

# B Proofs

## B.1 Proofs for Section 3

*Proof of Proposition 2.* ($\Rightarrow$) Let $D_r = (D_1, D_2, fg) \in D_m^{\pm,r}(M)$, we show that $(D_1, D_2) \in D_m^{\pm}(M)$.

We first show that $(D_1, D_2) \in D^{\pm}(M)$. Since $D_r$ is a refined diagnosis, it holds that $M[br(M) \setminus D_1 \cup fg(D_2)] \not\models \bot$. Let $S_w$ be a witnessing equilibrium of $M[br(M) \setminus D_1 \cup fg(D_2)]$, then it holds for every $r \in D_2$ that $S_w \models\!\!\leadsto fg(r)$ since $D_r$ is minimal. Therefore, $S_w$ is an equilibrium of $M[br(M) \setminus D_1 \cup cf(D_2)]$, hence $(D_1, D_2) \in D^{\pm}(M)$. Since $D_r$ is minimal, it holds for no $r \in D_2$ that $S_w \models\!\!\leadsto hd(r) \leftarrow body(fg(r)) \cup B$. where $body(fg(r)) \subset B \subseteq body(r)$, hence $\mathit{refine}(D_2, S_w) = fg$.

It remains to show that $(D_1, D_2) \in D^{\pm}_m(M)$. Assume for contradiction that there exists $(D'_1, D'_2) \subset (D_1, D_2)$ such that $(D'_1, D'_2) \in D^{\pm}_m(M)$. Let $S'_w$ be a witnessing equilibrium of $(D'_1, D'_2)$ and $fg' = \mathit{refine}(D'_2, S'_w)$, then it holds that $(D'_1, D'_2, fg') \in D^{\pm,r}(M)$ since $S'_w$ is a witnessing equilibrium of $M[br(M) \setminus D'_1 \cup fg'(D'_2)]$. Since $(D'_1, D'_2, fg') \in D^{\pm,r}(M)$ and $(D'_1, D'_2) \subset (D_1, D_2)$ it holds that $D_r$ is not a minimal refined diagnosis, which is a contradiction. Therefore, no such $(D'_1, D'_2)$ exists and $(D_1, D_2) \in D^{\pm}_m(M)$.

$(\Leftarrow)$ Let $D = (D_1, D_2) \in D^{\pm}_m(M)$, let $S_w$ be a witnessing equilibrium of $D$, and let $\mathit{refine}(D_2, S_w) = fg$. Furthermore, it holds that no witnessing equilibrium $S'_w$ exists with $\mathit{refine}(D_2, S'_w) = fg'$ and $fg' < fg$. We show that $(D_1, D_2, fg)$ is a minimal refined diagnosis of $M$. By definition of *refine* it holds for every $r \in D_2$ that $S_w \models\!\!\leadsto fg(r)$. Furthermore, there is no body-reduction function $fg'$ such that it holds for every $r \in D_2$ that $S_w \models\!\!\leadsto fg'(r)$. Therefore, $S_w$ is an equilibrium of $M[br(M) \setminus D_1 \cup fg(D_2)]$ and $(D_1, D_2, fg) \in D^{\pm,r}(M)$.

Towards contradiction assume that $(D_1, D_2, fg)$ is not minimal, then there exists $(D'_1, D'_2, fg') \in D^{\pm,r}_m(M)$ such that $(D'_1, D'_2) \subset (D_1, D_2)$ or $D_1 = D'_1$, $D_2 = D'_2$, and $fg' < fg$. In the former case, there exists a witnessing equilibrium $S'_w$ of $M[br(M) \setminus D'_1 \cup fg'(D'_2)]$. Therefore $S'_w$ is a witnessing equilibrium of $M[br(M) \setminus D'_1 \cup cf(\{r \in D'_2 \mid S'_w \models\!\!\leadsto fg'(r)\})]$, i.e., $D'' = (D'_1, \{r \in D'_2 \mid S'_w \models\!\!\leadsto fg'(r)\}) \in D^{\pm}(M)$. Since $D'' \subset D$ this is a contradiction to $D \in D^{\pm}_m(M)$. In the latter case holds $fg' < fg$ and there exists a witnessing equilibrium $S'_w$ of $M[br(M) \setminus D_1 \cup fg'(D_2)]$. Since $(D'_1, D'_2, fg') \in D^{\pm,r}_m(M)$ and $D'_1 = D_1, D'_2 = D_2$, it holds that $S'_w$ also is an equilibrium of $M[br(M) \setminus D_1 \cup cf(D_2)]$ and $\mathit{refine}(D_2, S'_w) = fg'$. Then, $fg' < fg$ directly contradicts that our assumption that no such $S'_w$ and $fg'$ exist. Since all cases are contradicting, it must hold that $(D_1, D_2, fg)$ is a minimal refined diagnosis. □

*Proof of Proposition 3.* $(\Rightarrow)$ Let $(E_1, E_2) \in E^{\pm}(M)$, pick $fg$ such that for every $r \in E_2$ holds $fg(r) = hd(r) \leftarrow .$, i.e. $\{fg(r) \mid r \in E_2\} = cf(E_2)$. Observe that for every $r \in E_2$ holds that $fg(r) = r$, therefore for all sets $R_1, R_2$ of bridge rules with $r \in R_2$ holds that $R_1 \cup fg(R_2) = R_1 \cup \{r\} \cup fg(R_2 \setminus \{r\})$. Then, for all $E_1 \subseteq R_1 \subseteq br(M)$, $R_2 \subseteq br(M)$, and body-reduction functions $fg'$ such that $body(fg(r)) \subseteq body(fg'(r))$ holds if $r \in E_2$, it holds that $M[R_1 \cup fg'(R_2)] \models \bot$, i.e., $(E_1, E_2, fg)$ is a refined explanation.
$(\Leftarrow)$ Let $(E_1, E_2, fg)$ be a refined explanation, i.e., $M[R_1 \cup fg'(R_2)] \models \bot$ for every $E_1 \subseteq R_1 \subseteq br(M)$, $R_2 \subseteq br(M)$, and body-reduction function $fg'$ with $body(fg(r)) \subseteq body(fg'(r))$ for every $r \in E_2$. Consider the body-reduction function $fg'$ such that for all $r \in br(M) \setminus E_2$ it holds that $fg'(r) = hd(r) \leftarrow .$, i.e., $fg'(R'_2) = cf(R'_2)$ for every $R'_2 \subseteq br(M) \setminus E_2$. Observe that $body(fg(r)) \subseteq body(fg'(r))$ holds for every $r \in E_2$, therefore $M[R_1 \cup fg'(R_2)] \models \bot$ for every $E_1 \subseteq R_1 \subseteq br(M)$ and $R_2 \subseteq br(M)$. Hence, $M[R_1 \cup cf(R'_2)] \models \bot$ for every $R'_2 \subseteq br(M) \setminus E_2$ and thus $(E_1, E_2) \in E^{\pm}(M)$. □

### B.2 Proofs for Section 4

*Proof of Theorem 1.* In this proof, for variables $E_i$, $D_i$, and $R_i$ with $i \in \{1, 2\}$, we assume that $E_i, D_i, R_i \subseteq br(M)$. Furthermore, we denote by $\overline{X}$ the complement of set $X$ wrt. $br(M)$, i.e., $\overline{X} = br(M) \setminus X$.

(a) Given a pair $(E_1, E_2)$. For all diagnoses $(D_1, D_2) \in D^{\pm}(M)$, $D_1 \cap E_1$ or $D_2 \cap E_2$ or both are nonempty iff

for all $(D_1, D_2)$ we have that
$$M[\overline{D_1} \cup cf(D_2)] \not\models \perp \text{ implies } D_1 \cap E_1 \neq \emptyset \text{ or } D_2 \cap E_2 \neq \emptyset$$

which (by reversing the implication and simplifying) is equivalent to

for all $(D_1, D_2)$ we have that
$$(D_1 \cap E_1 = \emptyset \text{ and } D_2 \cap E_2 = \emptyset) \text{ implies } M[\overline{D_1} \cup cf(D_2)] \models \perp.$$

As $A \cap B = \emptyset$ with $A, B \subseteq br(M)$ is equivalent to $A \subseteq \overline{B}$ we next obtain

for all $(D_1, D_2)$ we have that
$$(E_1 \subseteq \overline{D_1} \text{ and } D_2 \subseteq \overline{E_2}) \text{ implies } M[\overline{D_1} \cup cf(D_2)] \models \perp.$$

If we let $D_1 = \overline{R_1}$ and $D_2 = R_2$ this amounts to

for all $(R_1, R_2)$ we have that
$$(E_1 \subseteq R_1 \text{ and } R_2 \subseteq \overline{E_2}) \text{ implies } M[R_1 \cup cf(R_2)] \models \perp. \quad (35)$$

This proves the result (a) as this last condition is the one of an explanation $(E_1, E_2)$ in Definition 6. Note that, if $(\emptyset, \emptyset) \in D^{\pm}(M)$, then no explanation exists; this is intentional and corresponds to the definitions of diagnosis and explanation for consistent systems.

(b) As $minHS_M(X)$ contains the $\subseteq$-minimal elements in $HS_M(X)$, and $E^{\pm}_m(M)$ contains the $\subseteq$-minimal elements in $E^{\pm}(M)$, (b) follows from (a). □

*Proof of Corollary 1.* Let $min(X)$ be the set of $\subseteq$-minimal elements in a collection $X$ of sets. Then for every $(A, B) \in X \setminus min(X)$ there is a pair $(A', B') \in min(X)$ with $(A', B') \subseteq (A, B)$. Given $HS_M(min(X))$, every pair $(A, B) \in X \setminus min(X)$ is hit by every pair $(C, D) \in HS_M(min(X))$. Therefore $HS_M(min(X)) = HS_M(X)$. Then (a) immediately follows from Theorem 1 (a), and (b) immediately follows from Theorem 1 (b). □

*Proof of Lemma 1.* A collection of sets $C = \{C_1, \ldots, C_n\}$ over a universe, i.e., $C_i \subseteq U$, $1 \leq i \leq n$, can be seen as a *hypergraph* $\mathcal{H} = (U, C)$ with vertices $U$ and hyperedges $C_i \in C$. If no hyperedge $C_i$ is contained in any hyperedge $C_j$, $i \neq j$, it is called *simple*. A hitting set on $C$ is called *transversal*, and the hypergraph $(U, C')$ containing as hyperedges $C'$ all minimal hitting sets of the hypergraph $\mathcal{H}$ is called *transversal hypergraph* $Tr(\mathcal{H})$.

We can map a collection $X = \{X^1, \ldots, X^n\}$ of pairs $X^i = (X^i_1, X^i_2)$ of sets, $X^i_1$,$X^i_2 \subseteq U$ bijectively to a collection $\mu(X) = \{\mu(X^1), \ldots, \mu(X^n)\}$ over $U \cup \{u' \mid u \in U\}$ where $\mu(X^i_1, X^i_2) = X^i_1 \cup \{u' \mid u \in X^i_2\}$. Then, $(A, B)$ is a hitting set of $X$ iff $\mu(A, B)$ is a hitting set of $\mu(X)$, and well-known results for transversal hypergraphs [Berge, 1989] carry over to minimal hitting sets over pairs.

In particular, given a simple hypergraph $\mathcal{H} = \mu(X)$, it holds that $Tr(Tr(\mu(X))) = \mu(X)$. This directly translates into the lemma, because $\mu(X)$ is a simple hypergraph due to incomparability (also called the antichain property) of $X$, and $\mu$ is bijective, therefore transversal hypergraphs can be mapped back to minimal hitting sets. □

*Proof of Theorem 2.* From Corollary 1 (b) we have that $E_m^{\pm}(M) = minHS_M(D_m^{\pm}(M))$. Applying $minHS_M$ on both sides of this formula and then using Lemma 1 yields

$$minHS_M(E_m^{\pm}(M)) = minHS_M(minHS_M(D_m^{\pm}(M))) = D_m^{\pm}(M).$$

$\square$

*Proof of Proposition 5.* Let $(D_1, D_2) \in D_m^{\pm}(M)$ and let $S$ be a witnessing belief state for it, i.e., $S$ is an equilibrium of $M[br(M) \setminus D_1 \cup cf(D_2)]$. Towards contradiction, assume that $D_1 \cap D_2 \neq \emptyset$. Consider any bridge rule $r \in D_1 \cap D_2$ and let $h_c(r) = i$ and $h_b(r) = p$. Furthermore, consider $r' = cf(r) = (i : p) \leftarrow .$, then $body(r') = \emptyset$ and thus $r'$ is applicable in any belief state. Therefore, $r' \in app(br_i(M[br(M) \setminus D_1 \cup cf(D_2)]), S)$ and consequently $p \in \{h_b(r) \mid r \in app(br_i(M[br(M) \setminus D_1 \cup cf(D_2)]), S)\}$. For $(D'_1, D'_2) = (D_1 \setminus \{r\}, D_2)$, we thus obtain that $p \in \{h_b(r) \mid r \in app(br_i(M[br(M) \setminus D'_1 \cup cf(D'_2)]), S)\}$ and since all other bridge rules are as before, we conclude that $app(br_i(M[br(M) \setminus D'_1 \cup cf(D'_2)]), S) = app(br_i(M[br(M) \setminus D_1 \cup cf(D_2)]), S)$ for all $i \in c(M)$. Consequently $S$ is an equilibrium of $M[br(M) \setminus D'_1 \cup cf(D'_2)]$ and $(D'_1, D'_2) \in D^{\pm}(M)$. But $(D'_1, D'_2) \subset (D_1, D_2)$ contradicts $(D_1, D_2) \in D_m^{\pm}(M)$, which proves the result. $\square$

**Lemma 4.** *Let $U$ be a splitting set of an MCS $M$ and let $R_1, R_2 \subseteq br(M)$. Then, $U$ is also a splitting set of $M[R_1 \cup cf(R_2)]$.*

*Proof.* Towards contradiction assume that $U$ is not a splitting set for $M[R_1 \cup cf(R_2)]$, i.e., there exists a rule $r \in br(M[R_1 \cup cf(R_2)])$ such that $h_c(r) \in U$ and $b_c(r) \not\subseteq U$. Thus, there exists $(i : p) \in body(r)$ such that $i \notin U$. Since $body(r') = \emptyset$ for all $r' \in cf(R_2)$, it follows that $r \in R_1$ and since $R_1 \subseteq br(M)$, it follows that $r \in br(M)$. By the assumption that $h_c(r) \in U$ and because $U$ is a splitting set of $M$, it follows that $i \in U$ for all $(i : p) \in body(r)$, which contradicts that $b_c(r) \not\subseteq U$. Therefore, no such $r$ can exists and $U$ is also a splitting set of $M[R_1 \cup cf(R_2)]$. $\square$

**Lemma 5.** *Let $M$ be an MCS, let $B$ be a set of bridge rules compatible with $M$, and let $U \subseteq c(M)$ be a splitting set for $M[B]$. Then, for every $i \in U$ and belief state $S = (S_1, \ldots, S_n)$ of $M$ it holds that:*

$$S_i \in \mathbf{ACC}_i(kb_i \cup app(br_i(M[b_U]), S)) \text{ iff } S_i \in \mathbf{ACC}_i(kb_i \cup app(br_i(M[B]), S)).$$

*Proof.* We first show that $br_i(M[b_U]) = br_i(M[B])$ holds for all $i \in U$:

($\subseteq$) From the definition of the bottom, $b_U$, it follows that $b_U \subseteq B$, thus $br_i(M[b_U]) \subseteq br_i(M[B])$.

($\supseteq$) Consider $r \in br_i(M[B])$, it holds that $h_c(r) = i$. Since $U$ is a splitting set and $i \in U$ it follows that $r \in b_U$ by definition of the bottom $b_U$. Hence, $br_i(M[b_U]) \supseteq br_i(M[B])$.

As a consequence of the above, i.e., of $br_i(M[b_U]) = br_i(M[B])$, it follows that $app(br_i(M[b_U]), S) = app(br_i(M[B]), S)$ holds for all $i \in U$, and therefore it is also the case that

$$\mathbf{ACC}_i(kb_i \cup app(br_i(M[b_U]), S)) = \mathbf{ACC}_i(kb_i \cup app(br_i(M[B]), S)),$$

which proves the lemma. $\square$

Observe that splitting sets preserve acceptability not only when bridge rules in the remainder of the MCS are modified (as in Lemma 5), but also when belief sets in the remainder are exchanged. For two belief states $S = (S_1, \ldots, S_n)$ and $S' = (S'_1, \ldots, S'_n)$ of an MCS, we say that $S$ coincides with $S'$ on $U$, written $S =_U S'$, if for all $i \in U$ holds $S_i = S'_i$.

**Lemma 6.** *Let $M$ be an MCS, let $B$ be a set of bridge rules compatible with $M$, and let $U$ be a splitting set for $M[B]$. Furthermore, let $S = (S_1, \dots, S_n)$ and $S' = (S'_1, \dots, S'_n)$ be belief states of $M$, and let $b_U \subseteq R \subseteq B$. Then, $S =_U S'$ implies $\mathbf{ACC}_i(kb_i \cup app(br_i(M[B]), S,)) = \mathbf{ACC}_i(kb_i \cup app(br_i(M[R]), S',))$.*

*Proof.* Since $b_U \subseteq R$ it holds for all $i \in U$ that $br_i(M[B]) = br_i(M[R])$. Furthermore, because $U$ is a splitting set, it follows that $c \in U$ for all $(c : p) \in body(r)$ such that $r \in br_i(M[B])$ and $i \in U$. As a consequence $p \in S_c$ iff $p \in S'_c$ since $S$ and $S'$ coincide on $U$ and $r \in br_i(M[B])$ iff $r \in br_i(M[R])$. $\square$

*Proof of Proposition 6.* For reasoning about explanations, the concept of explanation range proves to be useful. For a given pair $E = (E_1, E_2) \in 2^{br(M)} \times 2^{br(M)}$ of sets of bridge rules and $B \subseteq br(M)$, the *explanation range* of $E$ with respect to $B$ is $Rg(E, B) = \{(R_1, R_2) \mid E_1 \subseteq R_1 \subseteq B \text{ and } R_2 \subseteq B \setminus E_2\}$. Intuitively, $Rg(E, B)$ are "relevant pairs" for $E$ wit respect to the upper bound $B$. It follows directly from Definition 6 that, $E = (E_1, E_2) \in E^{\pm}(M)$ iff $M[R_1 \cup cf(R_2)] \models \bot$ for all $(R_1, R_2) \in Rg(E, br(M))$.

In the following we prove Item (i): $E \in E^{\pm}(M[b_U])$ holds iff $E \in E^{\pm}(M)$ holds and $E$ is $U$-headed.

($\Rightarrow$) Let $(R'_1, R'_2) \in Rg(E, br(M))$ be arbitrary, then both $R'_1 \subseteq br(M)$ and $R'_2 \subseteq br(M)$. By Lemma 4, $U$ is also a splitting set for the MCS $N' = M[R'_1 \cup cf(R'_2)]$.

Let $R_1 = R'_1 \cap b_U$ and let $R_2 = R'_2 \cap b_U$. As $E_1, E_2 \subseteq b_U$, it follows that $(R_1, R_2) \in Rg(E, b_U)$. Because $E$ is an explanation of $M[b_U]$, it holds for $N = M[R_1 \cup cf(R_2)]$ that $N \models \bot$, i.e., for every belief state $S$ exists a context $i \in U$ with $S_i \notin \mathbf{ACC}_i(kb_i \cup app(br_i(N), S))$.

Since $B = R'_1 \cup cf(R'_2)$ is compatible with $M$ and $U$ is a splitting set for $N' = M[B]$, we conclude from Lemma 5 that for every belief state $S$ it holds that $S_i \in \mathbf{ACC}_i(kb_i \cup app(br_i(N'), S))$ iff $S_i \in \mathbf{ACC}_i(kb_i \cup app(br_i(N), S))$. Since $N \models \bot$ this implies that for every $S$ there exists some $i \in U$ such that $S_i \notin \mathbf{ACC}_i(kb_i \cup app(br_i(N'), S))$ and thus $N' \models \bot$.

Since $(R'_1, R'_2) \in Rg(E, br(M))$ is arbitrary, it follows that $E \in E^{\pm}(M)$. Furthermore, $E$ is $U$-headed by definition.

($\Leftarrow$) Let $E = (E_1, E_2) \in E^{\pm}(M)$ such that $E$ is $U$-headed, and consider some arbitrary $(R_1, R_2) \in Rg(E, b_U)$. Since $b_U \subseteq br(M)$, we conclude that $(R_1, R_2) \in Rg(E, br(M))$. Since $E$ is an explanation of $M$, it follows that $N = M[R_1 \cup cf(R_2)]$ is such that $N \models \bot$. As this holds for every $(R_1, R_2) \in Rg(E, b_U)$, it follows that $(E_1, E_2) \in E^{\pm}(M[b_U])$.

This establishes item (i).

Next we prove Item (ii): $D \in D^{\pm}(M[b_U])$ holds iff there exists $D' \in D^{\pm}(M)$ such that $D \subseteq D'$.

($\Rightarrow$) Let $D = (D_1, D_2) \in D^{\pm}(M[b_U])$. Then, there exists an equilibrium $S$ of $M[R]$ where $R = (b_U \setminus D_1) \cup cf(D_2)$. Consider $(D'_1, D'_2) = (D_1 \cup (br(M) \setminus b_U), D_2)$ and observe that $(br(M) \setminus D'_1) \cup cf(D'_2) = R$, because $br(M) \setminus D'_1 = b_U \setminus D_1$. Since $S$ is an equilibrium of $M[R]$, it follows that $D' \in D^{\pm}(M)$.

($\Leftarrow$) Assume $D' \in D^{\pm}(M)$ where $D' = (D'_1, D'_2)$. First assume that $E^{\pm}(M[b_U]) = \emptyset$, i.e., $M[b_U]$ is consistent. Then, $D = (\emptyset, \emptyset) \in D^{\pm}(M[b_U])$, hence $D \subseteq D'$ and $D \in D^{\pm}(M[b_U])$.

Otherwise, $E^{\pm}(M[b_U]) \neq \emptyset$. Consider $(D_1, D_2) = (D'_1 \cap b_U, D'_2 \cap b_U)$ and let $R' = br(M) \setminus D'_1 \cup cf(D'_2)$ and $R = b_U \setminus D_1 \cup cf(D_2)$. Observe that $br_j(M[R]) = \emptyset$ for all $j \in c(M) \setminus U$, because $R \subseteq b_U \cup cf(b_U)$ and for no rule $r \in b_U \cup cf(b_U)$ it holds that $h_c(r) = j$.

As $M[\emptyset]$ is consistent, there exists some $S^0_j \in \mathbf{ACC}_j(kb_j)$ for every $j \in c(M)$. Let $S' = (S'_1, \dots, S'_n)$ be an equilibrium for $M[R']$ (which exists because $D' \in D^{\pm}(M)$). Let $S = (S_1, \dots, S_n)$ such that $S_i = S'_i$ if $i \in U$, and $S_i = S^0_i$ otherwise. Then, $S$ is an equilibrium for $M[R]$. Indeed, first consider $i \in c(M) \setminus U$. Since $br_i(M[R]) = \emptyset$, it follows that $app(br_i(M[R]), S) = \emptyset$, hence $S^0_i \in \mathbf{ACC}_i(kb_i \cup app(br_i(M[R]), S))$. Second, consider $i \in U$. Note that $U$ is a splitting set of $M[R]$, because $br_j(M[R]) = \emptyset$ for all $j \in c(M) \setminus U$. Since $b_U \subseteq R \subseteq R'$ and $S =_U S'$, it follows from Lemma 6 that $\mathbf{ACC}_i(kb_i \cup$

$app(br_i(M[R]), S)) = \mathbf{ACC}_i(kb_i \cup app(br_i(M[R']), S'))$. From $S'_i \in \mathbf{ACC}_i(kb_i \cup app(br_i(M[R']), S'))$ and $S_i = S'_i$, it thus follows that $S_i \in \mathbf{ACC}_i(kb_i \cup app(br_i(M[R]), S))$.

Consequently, $S_i \in \mathbf{ACC}_i(kb_i \cup app(br_i(M[R]), S))$ for all $i \in c(M)$; hence $S$ is an equilibrium of $M[R]$. Since $R_1 \cup R_2 \subseteq b_U$, it follows that $D \in D^{\pm}(M[b_U])$. ☐

*Proof of Corollary 2.* Let $E \in E^{\pm}_m(M[b_U])$, then it follows from Proposition 6 that $E \in E^{\pm}(M)$ and $E$ is $U$-headed. Assume for a contradiction that $E \notin E^{\pm}_m(M)$. Hence, there exists some $E' \in E^{\pm}(M)$ such that $E' \subset E$. Since $E$ is $U$-headed, it follows that $E'$ also is $U$-headed. Thus by Proposition 6 it follows that $E' \in E^{\pm}(M[b_U])$, which contradicts that $E \in E^{\pm}_m(M[b_U])$. ☐

*Proof of Proposition 7.* As in the proof of Proposition 6, let $(R_1, R_2) \in Rg(E, B)$ iff $E_1 \subseteq R_1 \subseteq B$ and $R_2 \subseteq B \setminus E_2$.

Wlog. assume that $M = (C_1, \dots, C_n)$, $U = \{1, \dots, k\}$, and $U' = \{k+1, \dots, n\}$, where $1 \leq k \leq n$. Towards a contradiction assume that some $E = (E_1, E_2) \in E^{\pm}_m(M)$ exists which contains rules from both, $b_U$ and $b_{U'}$. Consider an arbitrary $(R_1, R_2) \in Rg(E, br(M))$. Since $E$ is an explanation, it holds that $M[R_1 \cup cf(R_2)] \models \bot$.

Given $R = (R_1, R_2)$ s.t. $R_1, R_2 \subseteq b_U \subseteq br(M)$ and $V$, we say that *the $V$-projection of $R$ is inconsistent* iff $M[(R_1 \cap V) \cup cf(R_2 \cap V)] \models \bot$. We prove that for every $R = (R_1, R_2) \in Rg(E, br(M))$ either its $U$-projection or its $U'$-projection is inconsistent, or both.

Towards contradiction assume that neither projection is inconsistent. Then, there exists an equilibrium $S = (S_1, \dots, S_n)$ of $M[(R_1 \cap U) \cup cf(R_2 \cap U)]$ and an equilibrium $S' = (S'_1, \dots, S'_n)$ of $M[(R_1 \cap U') \cup cf(R_2 \cap U')]$. Consider the belief state $S'' = (S_1, \dots, S_k, S'_{k+1}, \dots, S'_n)$. By Lemma 6, it holds that $S_i \in \mathbf{ACC}_i(kb_i \cup app(br_i(M[R_1 \cup cf(R_2)]), S'')$ for all $i \in U$, because $U$ is a splitting set of $M$, $b_U \subseteq (R_1 \cap U) \cup cf(R_2 \cap U) \subseteq R_1 \cup cf(R_2)$ , and $S =_U S''$. Analogously, it holds that $S'_i \in \mathbf{ACC}_i(kb_i \cup app(br_i(M[R_1 \cup cf(R_2)]), S'')$ for all $i \in U'$. Consequently, $S''$ is an equilibrium of $M[R_1 \cup cf(R_2)]$, which contradicts that $E$ is an explanation. Therefore, for every $R \in Rg(E, br(M))$ it holds that either the $U$-projection of $R$, the $U'$-projection of $R$, or both are inconsistent.

Next, we distinguish for all $R \in Rg(E, br(M))$ which projections are inconsistent.

Case (1): for every $R \in Rg(E, br(M))$ its $U$-projection is inconsistent. Then, $E' = (E_1 \cap b_U, E_2 \cap b_U)$ is an explanation, since for every $R' \in Rg(E', br(M))$ it holds that $R'$ is a $U$-projection of some $R \in Rg(E, br(M))$, which is inconsistent. Since $E_1 \cup E_2 \not\subseteq b_U$, we have $E' \subset E$. Since $E' \in E^{\pm}(M)$, it follows that $E \notin E^{\pm}_m(M)$, which contradicts the assumption that $E \in E^{\pm}_m(M)$.

Case (2): for all $R \in Rg(E, br(M))$ it holds that the $U'$-projection is inconsistent. Analogously to the previous case, we conclude that $E' = (E_1 \cap b_{U'}, E_2 \cap b_{U'})$ is an explanation of $M$ such that $E' \subset E$, which contradicts the assumption that $E \in E^{\pm}_m(M)$.

Case (3): Neither case (1) nor case (2) applies. That is, for some $R = (R_1, R_2) \in Rg(E, br(M))$ the $U$-projection is consistent, and also for some $R' = (R'_1, R'_2) \in Rg(E, br(M))$ the $U'$-projection is consistent. This means that there exists some belief state $S = (S_1, \dots, S_n)$ such that $S_i \in \mathbf{ACC}_i(kb_i \cup app(br_i(M[(R_1 \cap b_U) \cup cf(R_2 \cap b_U)]), S))$ for all $i \in c(M)$ and there exists some belief state $S' = (S'_1, \dots, S'_n)$ such that $S'_i \in \mathbf{ACC}_i(kb_i \cup app(br_i(M[(R'_1 \cap b_{U'}) \cup cf(R'_2 \cap b_{U'})]), S))$ for all $i \in c(M)$.

Now consider $R'' = (R''_1, R''_2) = ((R_1 \cap b_U) \cup (R'_1 \cap b_{U'}), (R_2 \cap b_U) \cup (R'_2 \cap b_{U'}))$. First, we show that $R'' \in Rg(E, br(M))$. Since $U$ and $U'$ partition $c(M)$ it holds that $E_1 = (E_1 \cap b_U) \cup (E_1 \cap b_{U'})$; since $E_1 \subseteq R_1$, clearly $E_1 \cap b_U \subseteq R_1 \cap b_U$. Analogously, it holds that $E_1 \cap b_{U'} \subseteq R_1 \cap b_{U'}$. Consequently, $E_1 = (E_1 \cap b_U) \cup (E_1 \cap b_{U'}) \subseteq (R_1 \cap b_U) \cup (R'_1 \cap b_{U'})$; hence $E_1 \subseteq R''_1 \subseteq br(M)$. For $R''_2$ observe that $(R_2 \cup R'_2) \cap E_2 = \emptyset$ since both, $R_2$ and $R'_2$, are disjoint with $E_2$ by definition. Therefore $((R_2 \cap b_U) \cup (R'_2 \cap b_{U'})) \cap E_2 = \emptyset$; hence $R''_2 \subseteq br(M) \setminus E_2$. In conclusion, it holds that $R'' \in Rg(E, br(M))$.

Second, we show that $S'' = (S_1, \ldots, S_k, S'_{k+1}, \ldots, S'_n)$ is an equilibrium of the MCS $M[R''_1 \cup cf(R''_2)]$. Since $S'' =_U S$ and as already shown, $R_1 \cap b_U \subseteq R''_1$ and $cf(R_2 \cap b_U) \subseteq cf(R''_2)$, it follows by Lemma 6 that $S_i \in \mathbf{ACC}_i(kb_i \cup app(br_i(M[R''_1 \cup cf(R''_2)]), S''))$ for all $i \in U$. Analogously, the same is shown for $U'$, i.e., $S'_i \in \mathbf{ACC}_i(kb_i \cup app(br_i(M[R''_1 \cup cf(R''_2)]), S''))$ for all $i \in U'$. Therefore, $S''$ is an equilibrium of $M[R''_1 \cup cf(R''_2)]$. Since $R'' \in Rg(E, br(M))$, it follows that $E \notin E^{\pm}(M)$. This is a contradiction to the assumption that $E \in E^{\pm}_m(M)$.

Since all cases yield a contradiction, it follows that every $E \in E^{\pm}_m(M)$ is either $U$-headed or $U'$-headed. □

*Proof of Corollary 3.* ($\subseteq$) Let $E^{\pm}_m(M)$. Then by Proposition 7, $E$ is either $U$-headed or $U'$-headed. If $E$ is $U$-headed, then by Proposition 6 $E \in E^{\pm}(M[b_U])$. Assume that $E \notin E^{\pm}_m(M[b_U])$, hence some $E' \subset E$ exists such that $E' \in E^{\pm}_m(M[b_U])$. By Proposition 7, $E' \in E^{\pm}_m(M)$. This contradicts that $E \in E^{\pm}_m(M)$, which gives $E \in E^{\pm}_m(M[b_U)$. Analogously, if $E$ is $U'$-headed, then $E \in E^{\pm}_m(M[b_{U'}])$. It follows that $E \in E^{\pm}_m(M[b_U]) \cup E^{\pm}_m(M[b_{U'}])$.

($\supseteq$) Let $E \in E^{\pm}_m(M[b_U])$ (respectively $E \in E^{\pm}_m(M[b_{U'}])$). Since $U$ (respectively $U'$) is a splitting set of $M$, from Corollary 2 it follows that $E \in E^{\pm}_m(M)$. In conclusion it holds that $E^{\pm}_m(M) \supseteq E^{\pm}_m(M[b_U]) \cup E^{\pm}_m(M[b_{U'}])$ □

*Proof of Proposition 8.* By Corollary 3, $E^{\pm}_m(M) = E^{\pm}_m(M[b_U]) \cup E^{\pm}_m(M[b_{U'}])$, while by Theorem 2 each diagnosis is a minimal hitting set on $E^{\pm}_m(M)$. Because $U$ and $U'$ partition $M$, $E^{\pm}_m(M[b_U])$ and $E^{\pm}_m(M[b_{U'}])$ are on disjoint sets. Therefore the minimal hitting set of their unions is the pairwise combination of their minimal hitting sets. That is, $(D_1, D_2) \in minHS_M(E^{\pm}_m(M))$ iff $(D_1, D_2) = (A_1 \cup B_1, A_2 \cup B_2)$ with $(A_1, A_2) \in minHS_M(E^{\pm}_m(M[b_U])$ and $(B_1, B_2) \in minHS_M(E^{\pm}_m(M[b_{U'}])$. From Theorem 2 it follows that $D^{\pm}_m(M) = minHS_M(M)$. This proves the proposition. □

### B.3 Proofs for Section 5

*Proof of Lemma 2.* ($\Rightarrow$) Let $S = (S_1, \ldots, S_n)$. Then $S_i \in \mathbf{ACC}(kb_i \cup H_i)$, where the set $H_i$ of active bridge rule heads at context $C_i$ is $app(br_i, S)$. Bridge rule applicability depends on output beliefs only, hence $app(br_i, S) = app(br_i, S^o)$. Thus $S^o = (S^o_1, \ldots, S^o_n)$ with $S^o_i = S_i \cap OUT_i$ is an output-projected equilibrium of $M$.

($\Leftarrow$) Let $S^o = (S^o_1, \ldots, S^o_n)$, then, as $S^o$ is an output-projected equilibrium, for each $i$, $1 \leq i \leq n$, $S^o_i \in \{T^o_i \mid T_i \in \mathbf{ACC}_i(kb_i \cup \{h_b(r) \mid r \in app(br_i, S^o)\})\}$, and therefore for each $S^o_i$ there exists some belief set $S_i$ such that $S_i \in \mathbf{ACC}_i(kb_i \cup \{h_b(r) \mid r \in app(br_i, S^o)\})$ and $S^o_i = S_i \cap OUT_i$. If we take for each $i$ some arbitrary $S_i$ satisfying the above condition, we obtain $T = (S_1, \ldots, S_n)$. As $S^o$ and $T$ agree on all output beliefs of all contexts, $app(br_i, S^o) = app(br_i, T)$ and hence $T$ is an equilibrium of $M$. By construction of $T$, it holds that also $T^o = S^o$. □

*Proof of Proposition 9.* (*Membership*) Given a MCS $M = (C_1, \ldots, C_n)$ we compute $OUT_i$ for all $C_i$ in $\mathcal{O}(|br(M)|)$, then we guess output projected belief sets $S^o_i \subseteq OUT_i$, $1 \leq i \leq n$, yielding an output-projected belief state $S^o$. We evaluate bridge rule applicability of all rules in $S^o$ in time $\mathcal{O}(|br(M)|)$ and thereby obtain a set of active bridge rule heads $H_i$ for each context $C_i$, $1 \leq i \leq n$. Finally we check acceptability of $S^o_i$ for all contexts $C_i$, i.e., whether $S^o_i \in \mathbf{ACC}_i(kb_i \cup H_i)|_{OUT_i}$. We accept if all contexts accept, otherwise we reject. This check is a conjunction of $n$ independent acceptability checks of maximum complexity equal to the smallest upper bound $C$ on context complexities which is $C = \mathcal{CC}(M)$. If $C$ is closed under conjunction, we can unite these checks into one check of complexity $C$ over an instance of size $\mathcal{O}(|M|)$. Then the overall acceptability check is in $C$ as well. This way we check the output-projected equilibrium property for all possible output-projected equilibria. Therefore if no computation path accepts, then the MCS $M$ is inconsistent. If there is one path that accepts, then the output-projected belief state $S^o$ corresponding to the guesses on this path is an output-projected equilibrium which fulfills all conditions of Definition 13. Therefore $M$ is consistent iff at least one path accepts. Hence if $C$ is closed under conjunction and projection, then the guess of size $\mathcal{O}(|br(M)|)$ can be projected away (i.e., incorporated into $I'$, see Section 5) and the complexity of MCSEQ is in $C$. For $\mathcal{CC}(M) = \mathbf{P}$ (which is presumably not closed under projection) the complexity of MCSEQ is in $\mathbf{NP}$.

($\mathbf{NP}$-*hardness* for $\mathcal{CC}(M) = \mathbf{P}$) We show that consistency checking in an MCS $M$ with lower context complexity $\mathcal{CC}(M) \geq \mathbf{P}$ is $\mathbf{NP}$-hard. We use the part of the MCS structure in Figure 3a labeled with MCSEQ. We reduce a 3-SAT instance $F = c_1 \wedge \ldots \wedge c_n$ on variables $\mathcal{X} = \{x_1, \ldots, x_k\}$ and clauses $c_i = c_{i,1} \vee c_{i,2} \vee c_{i,3}$ with $c_{i,j} \in \mathcal{X} \cup \{\neg x \mid x \in \mathcal{X}\}$ to consistency checking in an MCS $M = (C_{gen_U}, C_{eval_F}, C_{check})$. Context $C_{gen_U} = (L_{GUESS}, kb_{gen_U}, br_{gen_U})$ with $kb_{gen_U} = \mathcal{X}$ and $br_{gen_U} = \emptyset$ has linear complexity, while $C_{eval_F} = (L_{ASP}, kb_{eval_F}, br_{eval_F})$ and $C_{check} = (L_{ASP}, kb_{check}, br_{check})$ have context complexity $\mathbf{P}$. $M$ contains the following bridge rules:

$$r_{u,i}: \quad (eval_F : x_i) \leftarrow (gen_U : x_i). \qquad \forall i : 1 \leq i \leq k \tag{36}$$

$$r_\alpha: \quad (check : nsat) \leftarrow \mathbf{not}\ (eval_F : sat). \tag{37}$$

Hence $br_{eval_F} = \{r_{u,i} \mid \forall i : 1 \leq i \leq k\}$ and $br_{check} = \{r_\alpha\}$. The knowledge base $kb_{eval_F}$ is as follows:

$$sat_i \leftarrow l_{i,1}. \quad sat_i \leftarrow l_{i,2}. \quad sat_i \leftarrow l_{i,3}. \qquad \forall i : 1 \leq i \leq n \tag{38}$$

$$sat \leftarrow sat_1, \ldots, sat_n. \tag{39}$$

$$\text{where } l_{i,j} \text{ is } \begin{cases} x_v & \text{if } c_{i,j} = x_v \\ not\, x_v & \text{if } c_{i,j} = \neg x_v \end{cases}$$

The knowledge base $kb_{check}$ is as follows:

$$\bot \leftarrow nsat. \tag{40}$$

Context $C_{gen_U}$ accepts all possible subsets of $\mathcal{X}$, representing all possible truth assignments for the variables $\mathcal{X}$. (36) imports the truth assignment into $C_{eval_F}$, which evaluates $F$ under that truth assignment using rules (38) and (39). Then $C_{eval_F}$ puts the belief $sat$ in its belief set iff $F$ is satisfied given the truth assignment accepted by $C_{gen_U}$. Finally $C_{check}$ imports the belief $nsat$ iff $sat$ is not accepted at $C_{eval_F}$. Therefore constraint (40) makes $C_{check}$ inconsistent, i.e., accepts no belief set, iff $sat$ is not true in $C_{eval_F}$ iff there is no satisfying truth assignment for $F$. Therefore, if $F$ has a satisfying assignment with variables $\mathcal{T} \subseteq \mathcal{X}$ set to $\mathbf{t}$ and variables $\mathcal{X} \setminus \mathcal{T}$ set to $\mathbf{f}$, then $M$ has an equilibrium $S = (S_{gen_U}, S_{eval_F}, S_{check})$ where $S_{gen_U} = \mathcal{T}$, $S_{eval_F} = \mathcal{T} \cup \{sat_i \mid 1 \leq i \leq n\} \cup \{sat\}$, and $S_{check} = \emptyset$. Conversely, if $M$ has an equilibrium $S = (S_{gen_U}, S_{eval_F}, S_{check})$, then $S_{check}$ does not contain $nsat$ due to constraint (40). Hence $S_{eval_F}$ must contain $sat$, thus $S_{eval_F}$ contains $\{sat\} \cup \{sat_i \mid 1 \leq i \leq n\}$ due to (39). It follows that the set of bridge rule heads active at $C_{eval_F}$ corresponds to a satisfying assignment of $F$. This shows that MCS $M$ is consistent iff $F$ is a satisfiable 3-SAT instance. As the size of $M$ is linear in the size of the formula $F$ and 3-SAT is an $\mathbf{NP}$-hard problem, hardness for equilibrium existence follows.

($\mathcal{CC}(M)$-*hardness*) We show that consistency checking in an MCS $M$ with lower context complexity $\mathcal{CC}(M) \geq C$ is $C$-hard if $C$ is a class with complete problems that is closed under conjunction and projection. For that we use part of the MCS structure labeled with MCSEQ in Figure 3b. We reduce context acceptability checking, i.e., an instance $(H_a, S_a)$, $C_a = (kb_a, br_a, L_a)$ with $IN_a$, $OUT_a$ and context complexity $\mathcal{CC}(C_a)$ to consistency checking in an MCS $M = (C_{a'}, C_{check})$ such that the context complexity $\mathcal{CC}(C_{a'}) = \mathcal{CC}(C_a)$ and $\mathcal{CC}(C_{check}) = \mathbf{P}$. Intuitively, $C_{a'}$ gets input $H_a$, bridge rule $r_\alpha$ is applicable only if $S_a$ is accepted by $H_a$, and $C_{check}$ verifies whether $r_\alpha$ is applicable. Then $M$ is consistent iff $(H_a, S_a)$, $C_a$ is a 'yes' instance. Formally, $C_{a'} = (kb_a \cup H_a, \emptyset, L_a)$ uses knowledge base and logic from $C_a$, while $C_{check} = (kb_{check}, br_{check}, L_{ASP})$ use the specific logic $L_{ASP}$ that can be decided in $\mathbf{P}$. Bridge rules of $M$ are as follows:

$$r_\alpha: \quad (check : equal_{S'_a}) \leftarrow l_1, \dots, l_j, \dots l_{|OUT_a|}.$$

$$\text{where } l_j \text{ is } \begin{cases} s_j & \text{if } s_j \in OUT_a \wedge s_j \in S_a \\ \mathbf{not}\ s_j & \text{if } s_j \in OUT_a \wedge s_j \notin S_a \end{cases} \tag{41}$$

$$r_{en}: \quad (check : en) \leftarrow . \tag{42}$$

The knowledge base $kb_{check}$ is as follows:

$$\perp \leftarrow not\ equal_{S'_a}, en. \tag{43}$$

Bridge rule $r_{en}$ ensures that $C_{check}$ fulfills our assumption that a context without input is consistent. Wlog. we assume that $C_a$ accepts some belief set given input $H_a$. $C_{a'}$ contains the logic of $C_a$ and its knowledge base already contains bridge rule heads $H_a$. Therefore $C_{a'}$ accepts a belief set $S_a^{full}$, such that $S_a^{full} \cup OUT_a = S_a$, iff $(H_a, S_a)$, $C_a$ is a 'yes' instance. Therefore, belief state $S = (S_a^{full}, \{equal_{S'_a}, en\})$ is an equilibrium iff $(H_a, S_a)$, $C_a$ is a 'yes' instance. All belief states where $C_{a'}$ accepts a belief set $T$ with $T \cap OUT_a \neq S_a$ trigger constraint (43) and therefore lead to an inconsistency. Therefore $M$ has an equilibrium, and this equilibrium is $S$ iff context $(H_a, S_a)$, $C_a$ is a 'yes' instance for context acceptability checking. We thus have reduced context acceptability checking to consistency checking in $M$ and hardness follows.

(Alternative reduction for $\mathbf{NP}$-hardness with $\mathbf{P}$-contexts) Note that the above reduction for $\mathbf{P}$-contexts uses an acyclic MCS with stratified negation in bridge rules. Furthermore the context $C_{gen_U}$ accepts $2^{|\mathcal{X}|}$ belief sets and the contexts $C_{eval_F}$ and $C_{check}$ accept at most one belief set for any input. In the above reduction $\mathbf{NP}$-hardness arises from the nondeterminism of $C_{gen_U}$, i.e., from the number of belief sets potentially accepted by context $C_{gen_U}$. It is possible to obtain the hardness not from nondeterminism of a context but from nondeterminism of bridge rules. To illustrate this, we next give an alternative hardness reduction. (In subsequent proofs we only give one reduction, and there hardness arises from nondeterminism of contexts.) We reduce the same 3-SAT instance $F$ to an MCS $M = (C_1)$ consisting of one context $C_1 = (L_{ASP}, kb_1, br_1)$. It contains the following bridge rules $br_1$:

$$(1 : x_i) \leftarrow \mathbf{not}\ (1 : \bar{x}_i). \qquad \forall i : 1 \leq i \leq k \tag{44}$$

$$(1 : \bar{x}_i) \leftarrow \mathbf{not}\ (1 : x_i). \qquad \forall i : 1 \leq i \leq k \tag{45}$$

$$(1 : en) \leftarrow . \tag{46}$$

The knowledge base $kb_1$ is as follows:

$$sat_i \leftarrow l_{i,1}. \quad sat_i \leftarrow l_{i,2}. \quad sat_i \leftarrow l_{i,3}. \qquad \forall i : 1 \leq i \leq n \tag{47}$$

$$sat \leftarrow sat_1, \dots, sat_n. \tag{48}$$

$$\perp \leftarrow en, not\ sat. \tag{49}$$

$$\text{where } l_{i,j} \text{ is } \begin{cases} x_v & \text{if } c_{i,j} = x_v \\ \bar{x}_v & \text{if } c_{i,j} = \neg x_v \end{cases}$$

Without bridge rules, $en$ is not true in the knowledge base, hence the body of constraint (49) is never satisfied. Therefore $C_1$ satisfies our assumption that a context without bridge rules is consistent. The facts $x_i$ and $\bar{x}_i$ are contained only in heads of bridge rules (44) and (45) and not in heads of rules in $kb_1$. Furthermore bridge rules (44) and (45) are mutually exclusive in their applicability for each $1 \leq i \leq n$. Therefore these bridge rules guess for each $x_i$ whether $x_i$ or $\bar{x}_i$ is part of the set of facts added to $kb_1$. (47) and (48) evaluate $F$ wrt. the guess for $x_i$: if $x_i$ is added by a bridge rule, then $x_i = \mathbf{t}$ in $F$, otherwise $x_i = \mathbf{f}$. The value of $F$ wrt. the guess for $x_i$ and $\bar{x}_i$ is represented as $sat$ in $kb_1$. The constraint (49) makes the context inconsistent if $en$ is true and $sat$ is not true. Therefore if $F$ is satisfied with variables $\mathcal{T} \subseteq \mathcal{X}$ set to $\mathbf{t}$ and variables $\mathcal{X} \setminus \mathcal{T}$ set to $\mathbf{f}$, then $M$ has an equilibrium $(S_1)$ where $S_1 = \{x_i \mid x_i \in \mathcal{T}\} \cup \{\bar{x}_i \mid x_i \in \mathcal{X} \setminus \mathcal{T}\} \cup \{sat_i \mid 1 \leq i \leq n\} \cup \{sat, en\}$. Conversely, if $M$ has an equilibrium $(S_1)$, then $S_1$ contains $en$ due to the unconditional bridge rule (46). Hence $S_1$ must contain $sat$ due to constraint (49), and thus $S_1$ contains $\{sat_i \mid 1 \leq i \leq n\}$ due to (48). Therefore the guess of bridge rules (44) and (45) corresponds to a satisfying assignment of $F$. This shows that $M$ is consistent iff $F$ is satisfiable. Context $C_1$ uses logic $L_{ASP}$, therefore $\mathcal{CC}(M) \geq \mathcal{CC}(C_1) = \mathbf{P}$. As the size of $M$ is linear in the size of the formula $F$ and 3-SAT is an $\mathbf{NP}$-hard problem, hardness for equilibrium existence follows. □

*Proof of Proposition 10.* (*Membership*) Given MCS $M$ and $D_1, D_2 \subseteq br(M)$, we compute the modified MCS $M' = M[br(M) \setminus D_1 \cup cf(D_2)]$ and return the result of deciding MCSEQ on $M'$. By Definition 4, this returns 'yes' iff $(D_1, D_2) \in D^{\pm}(M)$. The transformation can be done in time $\mathcal{O}(|M|)$ therefore MCSD is in the same complexity class as MCSEQ.

(*Hardness*) Deciding whether $(\emptyset, \emptyset)$ is a diagnosis of $M$ can be decided by checking consistency of $M$, because $(\emptyset, \emptyset) \in D^{\pm}(M)$ iff $M$ is consistent. Therefore MCSD is as hard as MCSEQ for respective context complexity. □

*Proof of Proposition 11.* (*Membership*) Given an MCS $M$ and $D_1, D_2 \subseteq br(M)$, we solve two independent decision problems: (a) we decide whether $(D_1, D_2)$ is a diagnosis of $M$, and (b) we check whether a smaller diagnosis $(D', D'') \subset (D_1, D_2)$ exists in $M$. We return 'yes' if (a) returns 'yes' and (b) returns 'no'. Thus, this procedure returns 'yes' iff (a) $(D_1, D_2)$ is a diagnosis and (b) no $\subseteq$-smaller diagnosis exists. Therefore the computation yields the correct result. For (a) we decide MCSD on $M$ and $(D_1, D_2)$. For (b) we guess for each bridge rule in $D_1$ whether it is contained in $D'$, and for each bridge rule in $D_2$ whether it is contained in $D''$. Then we continue with the decision procedure MCSD on $M$ and $(D', D'')$, i.e., we guess presence of output belief sets, evaluate bridge rule applicability, and check acceptability for each context. Consequently for deciding (b) we decide the complement of a polynomial projection of MCSD. Therefore MCSD$_m$ is in the complexity class of solving the MCSD problem and independently solving the complement of a polynomially projected MCSD problem. Hence if $\mathcal{CC}(M)$ is closed under conjunction and projection, then the complexity of MCSD$_m$ is in $\mathcal{D}(\mathcal{CC}(M))$. For $\mathcal{CC}(M) = \mathbf{P}$ (which is presumably not closed under projection) the complexity of MCSD$_m$ is in $\mathbf{D_1^P}$.

($\mathbf{D^P}$-*hardness* for $\mathcal{CC}(M) = \mathbf{P}$) We reuse ideas from the MCSEQ hardness proof for 3-SAT, but we now use the complete topology shown in Figure 3a. We reduce two 3-SAT instances $F$ and $G$ on variables $\mathcal{X}$ and $\mathcal{Y}$, respectively, to minimal diagnosis recognition on MCS $M = (C_{gen_V}, C_{eval_F}, C_{gen_U}, C_{eval_G}, C_{check})$. Intuitively, $C_{gen_U}$ and $C_{eval_F}$ provide $\mathbf{NP}$-hardness for satisfiability of $F$, while $C_{gen_V}$ and $C_{eval_G}$ provide $\mathbf{coNP}$-hardness for unsatisfiability of $G$. $C_{gen_U}$ and $C_{eval_F}$ are constructed from $F$ exactly as for the proof of MCSEQ hardness. Similarly, $C_{gen_V}$ and $C_{eval_G}$ are constructed from $G$ with bridge rules $r_{v,j}$ transferring a guessed set $V \subseteq \mathcal{Y}$ from $C_{gen_V}$ to $C_{eval_G}$. The bridge rules in $M$ are as follows:

$$r_{u,i}: \quad (eval_F : x_i) \leftarrow (gen_U : x_i). \quad \forall i : 1 \leq i \leq |\mathcal{X}| \tag{50}$$

$$r_{v,j}: \quad (eval_G : y_j) \leftarrow (gen_V : y_j). \quad \forall j : 1 \leq j \leq |\mathcal{Y}| \tag{51}$$

$$r_\alpha: \quad (check : nsat_F) \leftarrow \mathbf{not}\ (eval_F : sat). \tag{52}$$

$$r_\gamma: \quad (check : nsat_G) \leftarrow \mathbf{not}\ (eval_G : sat). \tag{53}$$

Context $C_{check}$ has the following knowledge base $kb_{check}$:

$$\bot \leftarrow nsat_F. \tag{54}$$

$$\bot \leftarrow nsat_G. \tag{55}$$

If $F$ and $G$ are both satisfiable, $M$ is consistent so $(\emptyset, \emptyset) \in D^{\pm}_m(M)$. If $F$ is satisfiable and $G$ is unsatisfiable, $M$ is inconsistent and a minimal diagnosis for $M$ is $(\{r_\gamma\}, \emptyset)$. If $F$ and $G$ are both unsatisfiable, $M$ is inconsistent and a minimal diagnosis is $(\{r_\alpha, r_\gamma\}, \emptyset) \in D^{\pm}_m(M)$. If $F$ is unsatisfiable and $G$ is satisfiable, $M$ is inconsistent; $(\{r_\gamma\}, \emptyset)$ is no minimal diagnosis, because every diagnosis containing $r_\gamma$ in $D_1$ must also contain $r_\alpha$ in $D_1$ to restore consistency in $M$. Therefore $(\{r_\gamma\}, \emptyset)$ is a minimal diagnosis of $M$ iff $F$ is satisfiable and $G$ is unsatisfiable. Therefore recognizing a minimal diagnosis in an MCS with $\mathcal{CC}(M) = \mathbf{P}$ is hard for $\mathbf{D^P}$. Note that it is possible to do this reduction with one context that evaluates $F$ and $G$ and checks the result, using bridge rules that guess $U$ and $V$ and bridge rules that individually activate satisfiability checking for $F$ and $G$. However this would make the reduction less readable.

($\mathbf{D}(\mathcal{CC}(M))$-*hardness*) We show that recognizing minimal diagnoses in an MCS $M$ with lower context complexity $\mathcal{CC}(M) \geq C$ is hard for $\mathbf{D}(C$ if $C$ is a class with complete problems that is closed under conjunction and projection. We reduce two context complexity check instances $(H_a, S_a)$, $C_a$ with $IN_a$, $OUT_a$ and $(H_b, S_b)$, $C_b$ with $IN_b$, $OUT_b$ to an MCS $M = (C_{a'}, C_{b'}, C_{check})$ with the topology shown in Figure 3b. Similar to the generic hardness reduction for MCSEQ, we reduce $H_a$ and $C_a = (kb_a, br_a, L_a)$ to the context $C_{a'} = (kb_a \cup H_a, \emptyset, L_a)$ and we reduce $H_b$ and $C_b = (kb_b, br_b, L_b)$ to the context $C_{b'} = (kb_b \cup H_b, \emptyset, L_b)$. Then $\mathcal{CC}(C_{a'}) = \mathcal{CC}(C_a)$ and $\mathcal{CC}(C_{b'}) = \mathcal{CC}(C_b)$. Furthermore $C_{a'}$ accepts a belief set $S^{full}_a$ with $S^{full}_a \cap OUT_a = S_a$ iff $(H_a, S_a)$, $C_a$ is a 'yes' instance. Similarly $C_{b'}$ accepts a belief set $S^{full}_b$ with $S^{full}_b \cap OUT_b = S_b$ iff $(H_b, S_b)$, $C_b$ is a 'yes' instance. The bridge rules $br_{check}$ are as follows.

$$r_\alpha: \quad (check : equal_{S'_a}) \leftarrow l_1, \ldots, l_j, \ldots l_{|OUT_a|}. \quad \text{where } l_j \text{ is } \begin{cases} s_j & \text{if } s_j \in OUT_a \wedge s_j \in S_a \\ \mathbf{not}\ s_j & \text{if } s_j \in OUT_a \wedge s_j \notin S_a \end{cases} \tag{56}$$

$$r_\gamma: \quad (check : equal_{S'_b}) \leftarrow l_1, \ldots, l_j, \ldots l_{|OUT_b|}. \quad \text{where } l_j \text{ is } \begin{cases} s_j & \text{if } s_j \in OUT_b \wedge s_j \in S_b \\ \mathbf{not}\ s_j & \text{if } s_j \in OUT_b \wedge s_j \notin S_b \end{cases} \tag{57}$$

$$r_{en}: \quad (check : en) \leftarrow. \tag{58}$$

The knowledge base $kb_{check}$ is as follows:

$$n\_equal \leftarrow not\ equal_{S'_a}. \tag{59}$$

$$n\_equal \leftarrow not\ equal_{S'_b}. \tag{60}$$

$$\bot \leftarrow not\ n\_equal, en. \tag{61}$$

Bridge rule $r_{en}$ ensures that $C_{check}$ fulfills our assumption that a context without input is consistent. Wlog. we assume that $C_a$ and $C_b$ accept some belief set given input $H_a$ and $H_b$, respectively. Bridge rule $r_\alpha$ adds $equal_{S'_a}$ to $C_{check}$ iff the first instance $(H_a, S_a)$, $C_a$ we reduce from is a 'yes' instance. The same is true for $r_\gamma$, $equal_{S'_b}$ and the second instance. Therefore there exists an equilibrium $S = (S_a^{full}, S_b^{full}, \{equal_{S'_a}, equal_{S'_b}, en\})$ in $M$, i.e., $(\emptyset, \emptyset) \in D_m^{\pm}(M)$, iff both instances are 'yes' instances. Moreover, if the first instance is a 'yes' instance and the second instance is a 'no' instance, then the system is inconsistent and there is a minimal diagnosis $(\emptyset, \{r_\gamma\}) \in D_m^{\pm}(M)$. If both instances are 'no' instances, activating $equal_{S'_b}$ is not sufficient for restoring consistency, and a minimal diagnosis for $M$ is then $(\emptyset, \{r_\alpha, r_\gamma\})$. Therefore $(\emptyset, \{r_\gamma\})$ is a minimal diagnosis for $M$ iff $(H_a, S_a)$, $C_a$ is a 'yes' instance and $(H_b, S_b)$, $C_b$ is a 'no' instance of context acceptability checking. Therefore we have established that MCSD$_m$ is hard for $\mathbf{D}(C)$. Note that by nesting contexts $C_{a'}$ and $C_{b'}$ into a new context it is possible, although more complicated, to obtain a reduction with just one context that is hard for $\mathbf{D}(C)$. □

*Proof of Proposition 12.* (*Membership*) For deciding $(E_1, E_2) \in E^{\pm}(M)$, we guess $R_1, R_2 \subseteq br(M)$ and check whether $E_1 \subseteq R_2$ and $R_2 \subseteq br(M) \setminus E_2$. If not, we immediately reject, otherwise we decide MCSEQ of $M[R_1 \cup cf(R_2)]$. Then all execution paths reject iff $(E_1, E_2)$ is an explanation. Therefore, if $\mathcal{CC}(M)$ is a class with complete problems that is closed under conjunction and projection, the complexity is in **co**-$\mathcal{CC}(M)$. For $\mathcal{CC}(M) = \mathbf{P}$ (which is presumably not closed under projection) we obtain that MCSE is in **coNP**.

(**coNP**-*hardness* for $\mathcal{CC}(M) = \mathbf{P}$) We reuse the MCSEQ hardness proof where a 3-SAT instance $F$ was reduced to MCS $M = (C_{gen_U}, C_{eval_F}, C_{check})$. Then satisfiability of $F$ implies consistency, therefore $E^{\pm}(M) = \emptyset$, i.e., no inconsistency explanations exist. Unsatisfiability of $F$ implies inconsistency, and in that case, $(\{r_\alpha\}, \emptyset)$ is an inconsistency explanation of $M$. Therefore $(\{r_\alpha\}, \emptyset)$ is recognized as inconsistency explanation of $M$ iff $F$ is unsatisfiable. Therefore the problem MCSE in an MCS with $\mathcal{CC}(M) \geq \mathbf{P}$ is hard for **coNP**.

(**co**-$\mathcal{CC}(M)$-*hardness*) We reuse the MCSEQ hardness proof where we reduced an instance $I = (H_a, S_a)$, $C_a$ to a MCS $M_I = (C_{a'}, C_{check})$. If $I$ is a 'yes' instance, then $M_I$ is consistent so no inconsistency explanation exists. If $I$ is a 'no' instance, an inconsistency explanation of $M_I$ is $(\{r_{en}\}, \{r_\gamma\}) \in E^{\pm}(M_I)$. Therefore the problem MCSE in an MCS $M$ with lower context complexity $\mathcal{CC}(M) \geq C$ is hard for **co**-$C$ if $C$ is a class with complete problems that is closed under conjunction and projection. □

*Proof of Lemma 3.* We write $(A_1, A_2) \subset (B_1, B_2)$ iff both, $(A_1, A_2) \subseteq (B_1, B_2)$ and $(A_1, A_2) \neq (B_1, B_2)$.

($\Rightarrow$) Assume $Q = (Q_1, Q_2)$ is a minimal explanation. Contrary to the Lemma, assume there exists another explanation $Q'$, such that $Q' = (Q_1, Q_2 \setminus \{r\})$ with $r \in Q_2$ or $Q' = (Q_1 \setminus \{r\}, Q_2)$ with $r \in Q_1$. Then $Q' \subset Q$, therefore $Q$ is not minimal, contradicting the assumption.

($\Leftarrow$) Assume an explanation $Q = (Q_1, Q_2)$, and no pair $(Q_1, Q_2 \setminus \{r\})$ with $r \in Q_2$ or $(Q_1 \setminus \{r\}, Q_2)$ with $r \in Q_1$ is an explanation. Contrary to the Lemma, assume another explanation $P = (P_1, P_2)$ exists with $P \subset Q$. By $P \subset Q$, either a) $P_1 \subset Q_1$ and $P_2 \subseteq Q_2$ or b) $P_1 \subseteq Q_1$ and $P_2 \subset Q_2$. For a) we create $T' = (Q_1 \setminus \{r\}, Q_2)$ for some $r \in Q_1 \setminus P_1$. Then $P \subseteq T' \subset Q$. Due to Corollary 1, $T'$ is an explanation, contradicting the initial assumption. The case b) is similar. □

*Proof of Proposition 13.* (*Membership*) We can decide $(E_1, E_2) \in E_m^{\pm}(M)$ by using Lemma 3, i.e., we decide (1) whether $(E_1, E_2) \in E^{\pm}(M)$, and (2) whether all of $(E_1, E_2 \setminus \{r \mid r \in E_2\}) \notin E^{\pm}(M)$ and $(E_1 \setminus \{r \mid r \in E_1, E_2\}) \notin E^{\pm}(M)$ are true. Note that the number of $E^{\pm}$-checks in (2) is linear in the size of the instance, hence we obtain the following membership results: if the upper context complexity $\mathcal{CC}(M)$

is a class with complete problems that is closed under conjunction and projection, deciding (1) is in **co-**$\mathcal{CC}(M)$ and deciding (2) is in $\mathcal{CC}(M)$, therefore MCSE$_m$ is in $\mathcal{D}(\mathcal{CC}(M))$. For upper context complexity $\mathcal{CC}(M) = \mathbf{P}$ (which is not closed under projection) deciding (1) is in **coNP** and deciding (2) is in **NP** and therefore MCSE$_m$ is in $\mathbf{D_1^P}$.

($\mathbf{D^P}$*-hardness* for $\mathcal{CC}(M) = \mathbf{P}$) We use the same topology as for the MCSD$_m$ hardness proof, i.e., the complete topology shown in Figure 3a. We reduce two 3-SAT instances $F$ and $G$ on variables $\mathcal{X}$ and $\mathcal{Y}$, respectively, to minimal explanation recognition on MCS $M = (C_{gen_V}, C_{eval_F}, C_{gen_U}, C_{eval_G}, C_{check})$. Again, $C_{gen_U}$ and $C_{eval_F}$ provide **NP**-hardness for satisfiability of $F$, while $C_{gen_V}$ and $C_{eval_G}$ provide **coNP**-hardness for unsatisfiability of $G$. All contexts except for $C_{check}$ are constructed from $F$ and $G$ exactly as in the MCSD$_m$ hardness proof. Wlog. we assume that $F$ is not valid. The bridge rules in $M$ are as follows:

$$r_{u,i}: \quad (eval_F : x_i) \leftarrow (gen_U : x_i). \qquad \forall i : 1 \leq i \leq |\mathcal{X}| \tag{62}$$

$$r_{v,j}: \quad (eval_G : y_j) \leftarrow (gen_V : y_j). \qquad \forall j : 1 \leq j \leq |\mathcal{Y}| \tag{63}$$

$$r_\alpha: \quad (check : sat\_or\_nsat_F) \leftarrow (eval_F : sat). \tag{64}$$

$$r_\beta: \quad (check : sat\_or\_nsat_F) \leftarrow \mathbf{not}\ (eval_F : sat). \tag{65}$$

$$r_\gamma: \quad (check : nsat_G) \leftarrow \mathbf{not}\ (eval_G : sat). \tag{66}$$

Context $C_{check}$ has the following knowledge base $kb_{check}$:

$$\bot \leftarrow sat\_or\_nsat_F, not\ nsat_G. \tag{67}$$

If $G$ is satisfiable, $M$ is consistent, so $E_m^{\pm}(M) = \emptyset$. If $F$ and $G$ are unsatisfiable, the belief $sat$ is never accepted at $C_{eval_F}$; therefore the bridge rule $r_\beta$ is sufficient for creating inconsistency in $M$ (i.e., $M[\{r_\beta\}] \models \bot$) and forcing $r_\gamma$ to become applicable is the only way to restore consistency. Therefore if $F$ and $G$ are unsatisfiable, $(\{r_\beta\}, \{r_\gamma\}) \in E^{\pm}(M)$. If $F$ is satisfiable and $G$ is unsatisfiable, the belief $sat$ may or may not be accepted at $C_{eval_F}$, depending on the input $C_{eval_F}$ gets from $C_{gen_U}$. Therefore both bridge rules $r_\alpha$ and $r_\beta$ are required for ensuring inconsistency in $M$, and they are also sufficient. Again, forcing $r_\gamma$ to become applicable is the only way to restore consistency. Therefore if $F$ is satisfiable and $G$ is unsatisfiable, then $(\{r_\alpha, r_\beta\}, \{r_\gamma\}) \in E^{\pm}(M)$. Thus $(\{r_\alpha, r_\beta\}, \{r_\gamma\})$ is a minimal inconsistency explanation for $M$ iff $F$ is satisfiable and $G$ is unsatisfiable. Note that if $G$ is satisfiable, no explanations exist, while if $F$ is unsatisfiable, the above explanation exists but is no longer minimal. Therefore recognizing a minimal inconsistency explanation in an MCS with $\mathcal{CC}(M) = \mathbf{P}$ is hard for $\mathbf{D^P}$.

($\mathbf{D}(\mathcal{CC}(M))$*-hardness*) We use the same topology as for the MCSD$_m$ hardness proof, i.e., the complete topology shown in Figure 3b. We also use a very similar reduction. The only change is in the checking context $C_{check}$. We reduce two context complexity check instances $(H_a, S_a)$, $C_a$ with $IN_a$, $OUT_a$ and

$(H_b, S_b)$, $C_b$ with $IN_b$, $OUT_b$ to an MCS $M = (C_{a'}, C_{b'}, C_{check})$. The bridge rules $br_{check}$ are as follows.

$$r_\alpha: \quad (check : equal_{S'_a}) \leftarrow l_1, \ldots, l_j, \ldots l_{|OUT_a|}. \quad \text{where } l_j \text{ is } \begin{cases} s_j & \text{if } s_j \in OUT_a \wedge s_j \in S_a \\ \mathbf{not}\ s_j & \text{if } s_j \in OUT_a \wedge s_j \notin S_a \end{cases} \tag{68}$$

$$r_\beta: \quad (check : make\_inc) \leftarrow l_1, \ldots, l_j, \ldots l_{|OUT_a|}. \quad \text{where } l_j \text{ is } \begin{cases} s_j & \text{if } s_j \in OUT_a \wedge s_j \in S_a \\ \mathbf{not}\ s_j & \text{if } s_j \in OUT_a \wedge s_j \notin S_a \end{cases} \tag{69}$$

$$r_\gamma: \quad (check : equal_{S'_b}) \leftarrow l_1, \ldots, l_j, \ldots l_{|OUT_b|}. \quad \text{where } l_j \text{ is } \begin{cases} s_j & \text{if } s_j \in OUT_b \wedge s_j \in S_b \\ \mathbf{not}\ s_j & \text{if } s_j \in OUT_b \wedge s_j \notin S_b \end{cases} \tag{70}$$

$$r_{en}: \quad (check : en) \leftarrow . \tag{71}$$

Note that $r_\alpha$ and $r_\beta$ have the same body but different heads, moreover only $r_\beta$ differs from the MCSD$_m$-reduction. The knowledge base $kb_{check}$ is as follows:

$$n\_equal_a \leftarrow not\ equal_{S'_a}. \tag{72}$$

$$n\_equal_a \leftarrow make\_inc. \tag{73}$$

$$\bot \leftarrow en, n\_equal_a, not\ equal_{S'_b}. \tag{74}$$

The bridge rule $r_{en}$ ensures that $C_{check}$ fulfills our assumption that a context without input is consistent. Wlog. we assume that $C_a$ and $C_b$ accept some belief set given input $H_a$ and $H_b$, respectively. The bridge rule $r_\alpha$ adds $equal_{S'_a}$ to $C_{check}$ iff the first instance $(H_a, S_a)$, $C_a$ is a 'yes' instance. Under the same condition, $r_\beta$ adds $make\_inc$. The bridge rule $r_\gamma$ adds $equal_{S'_b}$ to $C_{check}$ iff the second instance $(H_b, S_b)$, $C_b$ is a 'yes' instance. In that case, $M$ is consistent, i.e., $E^{\pm}_m(M) = \emptyset$, because $r_\gamma$ becomes applicable and this deactivates constraint (74) such that $C_{check}$ can no longer become inconsistent. If both instances are 'no' instances, $M$ is inconsistent and for explaining this inconsistency it is sufficient to have $r_{en}$ present and the heads of the bridge rules $r_\alpha$ and $r_\gamma$ absent. Therefore, in that case, $(\{r_{en}\}, \{r_\alpha, r_\gamma\})$ is a minimal inconsistency explanation for $M$. Finally, if $(H_a, S_a)$, $C_a$ is a 'yes' instance and $(H_b, S_b)$, $C_b$ is a 'no' instance, $M$ is inconsistent and for this inconsistency it is sufficient to have $r_{en}$ and $r_\beta$ present and heads of bridge rules $r_\alpha$ and $r_\gamma$ absent, so $(\{r_{en}, r_\beta\}, \{r_\alpha, r_\gamma\}) \in E^{\pm}_m(M)$. Therefore $(\{r_{en}, r_\beta\}, \{r_\alpha, r_\gamma\}) \in E^{\pm}_m(M)$ iff the first instance is a 'yes' instance and the second instance is a 'no' instance. Note that if the second instance is a 'yes' instance, no explanations exist, while if the first instance is a 'no' instance, the above explanation exists but is no longer minimal. Therefore we have established that MCSE$_m$ is hard for $\mathbf{D}(C)$ where $\mathcal{CC}(M) = C$ if $C$ is a class with complete problems that is closed under conjunction and projection. □

### B.4 Proofs for Section 6

**Preliminaries for proving the results of Section 6** For proving the correctness of our HEX encodings, we use some lemmas.

**Lemma 7.** *Let $P = R \cup C$ be a* HEX *program consisting of an ordinary* HEX*-program $R$ and a set of constraints $C$ which contain external atoms. Then for every $I \in \mathcal{AS}(P)$ it holds that $I \in \mathcal{AS}(R)$ and $I$ does not satisfy the body of any constraint in $C$.*

*Proof.* From $I \in \mathcal{AS}(P)$ we know that $I \models P$ and therefore $I \models R$ and $I \models C$. From the latter we infer that $I$ does not satisfy the body of any constraint in $C$ (i.e., the second claim). Thus the reduct $fP^I$ does not contain any constraint from $C$. Hence $fP^I = fR^I$ and $I$ is a minimal model of $fR^I$ as it is a minimal model of $fP^I$. □

**Lemma 8.** *Let $P$ be a* HEX *program, and let $I \in \mathcal{AS}(P)$ be an answer set of $P$. Then for every atom $a \in I$ it holds that there is a rule $r \in P$ of form* (2) *with $a \in \{a_1, \ldots, a_k\}$ and $I$ satisfies the body of $r$.*

*Proof.* Assume towards a contradiction that $I \in \mathcal{AS}(P)$, $a \in I$ and no rule $r \in P$ is such that $a$ is in the head of $r$ and $I$ satisfies the body of $r$. Due to the latter assumption, no rule that contains $a$ in the head is contained in $fP^I$. Since $I$ is an answer set of $P$, $I \models fP^I$, therefore the bodies of all rules in $fP^I$ are satisfied by $I$. Hence every rule in $fP^I$ has a nonempty intersection of its head with $I$ (otherwise $I \not\models fP^I$). Because no rule in $fP^I$ contains $a$ in the head, it follows that $I \setminus \{a\} \models fP^I$, therefore $I$ is no minimal model of $fP^I$ and no answer set, which is a contradiction. □

*Proof of Theorem 4.* ($\Rightarrow$) Given $I \in \mathcal{AS}(P_D(M))$, due to Lemma 7 we have (a) $I \in AS(R)$ where $R$ contains rules (3), and (b) constraint (5) has an unsatisfied body. Due to (a) the pair $(D_{I,1}, D_{I,2})$ is such that $D_{I,1}, D_{I,2} \subseteq br(M)$. From (b) we know that the external atoms in (5) evaluate to true, therefore from (4) we know $M[br(M) \setminus D_{I,1} \cup cf(D_{I,2})] \not\models \bot$, hence $(D_{I,1}, D_{I,2}) \in D^{\pm}(M)$.

($\Leftarrow$) Given $(D_1, D_2) \in D^{\pm}(M)$ with $D_1 \cap D_2 = \emptyset$, the corresponding $Q = \{um(r) \mid r \in br(M) \setminus (D_1 \cup D_2)\} \cup \{d_1(r) \mid r \in D_1\} \cup \{d_2(r) \mid r \in D_2\}$ satisfies rules (3), furthermore $P_D(M)$ contains only constraint (5) apart from (3), and this constraint, per definition of $f_{\&eq_M}$, has an unsatisfied body if $(D_1, D_2) \in D^{\pm}(M)$. Therefore the reduct $fP_D(M)^Q$ contains only the rules (3). For each rule $r \in br(M)$, $Q$ contains exactly one atom from the set $\{um(r), d_1(r), d_2(r)\}$. Hence $Q$ satisfies the reduct $fP_D(M)^Q$, furthermore for each atom we remove from $Q$, $Q$ no longer satisfy one rule in (3). Therefore $Q$ is a minimal model of $fP_D(M)^Q$ and hence $Q \in \mathcal{AS}(P_D(M))$. □

*Proof of Proposition 14.* (i) Given $I \in \mathcal{AS}(P_p(M))$, due to Lemma 7 we have (a) $I \in \mathcal{AS}(R)$ where $R$ contains rules (6), (7), and (8); and (b) no constraint (9) has a satisfied body. In $R$, (6) are the only rules with $pres_i$ and $abs_i$ atoms in the head, therefore $A_i(I) \subseteq OUT_i$ for each context $C_i \in c(M)$. Hence $\mathcal{A}(I)$ is an output-projected belief state of $M$. Due to Lemma 8, $I$ does not contain $d_1(r)$ or $d_2(r)$ for any $r \in br(M)$, as no rule contains these atoms in the head; therefore (8) never has a satisfied body and $I$ always satisfies $not\, d_1(r)$ in (7). Due to Lemma 8, $I$ contains $b_i(s)$ iff there is at least one bridge rule $r \in br(M)$ such that in the corresponding rule (7), for all $i$, $1 \leq i \leq j$, $pres_{c_i}(p_i) \in I$, and for all $l$, $j < l \leq m$, $pres_{c_l}(p_l) \notin I$. This in turn is the case iff for all $(c_i : p_i)$ in the body of $r$, $p_i \in A_{c_i}(I)$, and for all $\mathbf{not}$ $(c_l : p_l)$ in the body of $r$, $p_l \notin A_{c_l}(I)$. The same is true iff bridge rule $r$ is applicable in $\mathcal{A}(I)$, therefore we have $B_i(I) = \{h_b(r) \mid r \in app(br_i, \mathcal{A}(I))\}$ for each $C_i \in c(M)$. From (b) we can infer that for every context $C_i \in c(M)$, constraint (9) has an unsatisfied body, therefore the external atom returns false, hence $A_i(I) \in \mathbf{ACC}_i(kb_i \cup B_i(I))|_{OUT_i}$. We further obtain $A_i(I) \in \mathbf{ACC}_i(kb_i \cup \{h_b(r) \mid r \in app(br_i, \mathcal{A}(I))\})|_{OUT_i}$ for every $C_i \in c(M)$, which exactly satisfies Definition 13. Therefore $\mathcal{A}(I)$ is an output-projected equilibrium of MCS $M$.

(ii) Given an output-projected equilibrium $S^o = (S^o_1, \ldots, S^o_n)$ of $M$, we assemble an interpretation $I$ of $P_p(M)$ as follows: $I = \{a_i(p) \mid p \in S^o_i,\ 1 \leq i \leq n\} \cup \{\bar{a}_i(p) \mid p \in OUT_i \setminus S^o_i,\ 1 \leq i \leq n\} \cup \{b_i(s) \mid s \in H_i,\ 1 \leq i \leq n\}$, with $H_i = app(br_i, S^o)$. Facts (6) are contained in the reduct $fP_p(M)^I$. By construction of $I$ and by the definition of bridge rule applicability, and because $d_1$ has an empty extension in $I$, all bodies of rules (7) which correspond to an applicable bridge rule in $S^o$ are satisfied, therefore these rules are part of $fP_p(M)^I$. Because $d_2$ has an empty extension in $I$, no rule from (8) is part of $fP_p(M)^I$. Since $S^o$ is an output-projected equilibrium, for each $C_i$ it holds that $S_i^o \in \{T_i^o \mid T_i \in \mathbf{ACC}_i(kb_i \cup H_i)\}$. As $B_i(I) = H_i$ and $A_i(I) = S_i^o$, we obtain that $A_i(I) \in \mathbf{ACC}_i(kb_i \cup B_i(I))|_{OUT_i}$, therefore $f_{\&con\_out_i}(I, a_i, b_i) = 1$ for all $C_i$, and $I$ does not satisfy the body of any constraint (9). Hence none of the constraints (9) is part of $fP_p(M)^I$. $I$ satisfies all rules of $P_p(M)$ and all rules of $fP_p(M)^I$. Moreover, $P_p(M)$ does not contain loops, neither does $fP_p(M)^I$, hence $I$ is a $\subseteq$-minimal model of $fP_p(M)^I$ and therefore $I \in \mathcal{AS}(P_p(M))$. □

*Proof of Theorem 5.* (i) Given $I \in \mathcal{AS}(P_p^D(M))$, due to Lemma 7 we have (a) $I \in \mathcal{AS}(R)$ where $R$ contains rules (3), (6), (7), and (8); and (b) no constraint (9) has a satisfied body. As in $P_p(M)$, every $I$ corresponds to a unique belief state $\mathcal{A}(I)$ of $M$, and as in $P^D(M)$, every $I$ corresponds to a unique pair $(D_{I,1}, D_{I,2})$, $D_{I,1}$, $D_{I,2} \subseteq br(M)$. Due to Lemma 8, $I$ contains $b_i(s)$ iff at least one of the following is true: $d_2(r) \in I$ and accordingly $r \in D_{I,2}$, or there is at least one bridge rule $r \in br(M)$ such that $d_1(r) \notin I$ and in the corresponding rule (7) we have that for all $i$, $1 \leq i \leq j$, $pres_{c_i}(p_i) \in I$, and for all $l$, $j < l \leq m$, $pres_{c_l}(p_l) \notin I$; this holds iff $r \notin D_{I,1}$ and $r \in app(br_i(M), \mathcal{A}(I))$, which holds iff $r \in app(br_i(M) \setminus D_{I,1}, \mathcal{A}(I))$. Therefore, for each context $C_i \in c(M)$ we have $B_i(I) = \{h_b(r) \mid r \in app(br_i(M) \setminus D_{I,1}, \mathcal{A}(I))\} \cup \{h_b(r) \mid r \in D_{I,2}\}$. The condition-free bridge rules are always applicable, therefore $B_i(I) = \{h_b(r) \mid r \in app(br_i(M[br(M) \setminus D_{I,1} \cup cf(D_{I,2})]), \mathcal{A}(I))\}$. Note that in this expression, first all bridge rules of $M$ are modified using $D_{I,1}$ and $D_{I,2}$, then the bridge rules of context $C_i$ of the result are extracted using $br_i(\cdot)$. From (b) we know that for every context $C_i \in c(M)$, the external atom in (9) returns false, therefore $A_i(I) \in \mathbf{ACC}_i(kb_i \cup B_i(I))|_{OUT_i}$ for every $C_i \in c(M)$. Substituting $B_i(I)$ we obtain $A_i(I) \in \mathbf{ACC}_i(kb_i \cup \{h_b(r) \mid r \in app(br_i(M[br(M) \setminus D_{I,1} \cup cf(D_{I,2})]), \mathcal{A}(I))\})|_{OUT_i}$. Therefore $\mathcal{A}(I)$ is an output-projected equilibrium of MCS $M[br(M) \setminus D_{I,1} \cup cf(D_{I,2})]$ and $(D_{I,1}, D_{I,2}) \in D^{\pm}(M)$.

(ii) Given a diagnosis $(D_1, D_2) \in D^{\pm}(M)$ and given an output-projected equilibrium $S^o = (S_1^o, \ldots, S_n^o)$ of $M' = M[br(M) \setminus D_1 \cup cf(D_2)]$ we assemble the interpretation

$$I = \{d_1(r) \mid r \in D_1\} \cup \{d_2(r) \mid r \in D_2\} \cup \{um(r) \mid r \notin (D_1 \cup D_2)\} \cup \\ \{a_i(p) \mid p \in S_i^o\} \cup \{\bar{a}_i(p) \mid p \in OUT_i \setminus S_i^o\} \cup \{b_i(s) \mid s \in H_i\}$$

where $H_i = app(br_i(M'), S^o)$. Since $S^o$ is an output-projected equilibrium, $I$ satisfies constraints (9), therefore they are not part of the reduct $fP_p^D(M)^I$. By construction of $I$, those rules in (7) where $r \in D_1$ or $r$ is not applicable in $\mathcal{A}(I)$ have an unsatisfied rule body, so these rules are not part of the reduct. Those rules in (8) where $r \in D_2$ have a satisfied rule body, so these rules are always part of the reduct. Other rules in (7) or (8) are satisfied by $I$ as their body is not satisfied. For each applicable bridge rule $r$, the according head atom $b_i(s)$ is part of $I$, and $P_p^D(M)$ contains no cyclic dependencies between rules (hence neither does the reduct $fP_p^D(M)^I$). Therefore $I$ is a minimal model of rules (7) and (8) in the reduct. Rules (3) and (6) are contained in the reduct, and $I$ by construction is a minimal model of these rules. Therefore, $I$ is a model of $P_p^D(M)$ and a minimal model of $fP_p^D(M)^I$, hence $I \in \mathcal{AS}(P_p^D(M))$. □

**Preliminaries for the proof of Theorem 6** For the following proofs we assume $M = (C_1, \ldots, C_n)$ to be an arbitrary but fixed MCS and $P_P^E(M)$ to be the explanation encoding for $M$.

Given a HEX rule $r$ of form (2), we write $B_{hex}(r) = \{\beta_1, \ldots, \beta_n\}$ and $H_{hex}(r) = \{\alpha_1, \ldots, \alpha_k\}$ to denote body and head of $r$ respectively. For an interpretation $I$ and a HEX rule $r$, we write $I \models B_{hex}(r)$ iff $I \models \beta_i$ for all $i \in \{1, \ldots, m\}$ and $I \not\models \beta_j$ for all $j \in \{m+1, \ldots, n\}$. Similarly, we write $I \models H_{hex}(r)$ iff $I \models \alpha_i$ for some $i \in \{1, \ldots, k\}$.

For referring to a specific rule of $P_P^E(M)$, we write $tr_N(v_1, \ldots, v_\ell)$ where $N$ is the rule of form $(N)$ instantiated with $v_1 \ldots, v_\ell$. We denote by $TR_n(M)$ the set of all instantiations of a rule wrt. an MCS $M$. For example, let $r_7 \in br(M)$, then $tr_{12}(r_7)$ denotes the HEX rule $r1(r_7) :- e1(r_7).$, while $TR_{12}(M) = \{tr_{12}(r) \mid r \in br(M)\}$. For brevity, we write only those values necessary to identify the instantiation, e.g., for rules of form (17) we write $tr_{17}(r)$ where $r \in br(M)$; for a rule of form (25), we write $tr_{25}(i, b)$ where $(i : b)$ is the head of some $r \in br(M)$.

We say an interpretation $I$ *consistently encodes* an explanation candidate $E = (E_1, E_2)$ where $E_1 = \{r \in br(M) \mid e1(r) \in I\}$, $E_2 = \{r \in br(M) \mid e2(r) \in I\}$, for all $r \in br(M)$: (i) $e1(r) \in I$ iff $ne1(r) \notin I$, and (ii) $e2(r) \in I$ iff $ne2(r) \notin I$.

**Lemma 9.** *Every answer set $I$ of $P_P^E(M)$ consistently encodes an explanation candidate.*

*Proof.* Let $I$ be an answer set of $P_P^E(M)$. Then, by definition $I$ must be a minimal model of $fP_P^E(M)^I$. Assume for contradiction that $I$ does not consistently encode an explanation candidate. Then, for some $r \in br(M)$ one of the following cases holds.

(i) $e1(r) \in I$ and $ne1(r) \in I$: Consider $I' = I \setminus \{e1(r)\}$. For all $tr \in fP_P^E(M)^I$ with $e1(r) \notin H_{hex}(tr)$ it holds that $I' \models tr$ since $I \models tr$. There is only one rule $tr'$ such that $e1(r) \in H_{hex}(tr')$, namely $tr' = tr_{10}(r)$. Since $ne1(r) \in I'$ and $ne1(r) \in H_{hex}(tr_{10}(r))$ it holds that $I' \models tr$, hence $I' \models fP_P^E(M)^I$. Since $I' \subset I$ this contradicts that $I$ is a minimal model of $fP_P^E(M)^I$.

(ii) $e1(r) \notin I$ and $ne1(r) \notin I$. Since $B_{hex}(tr_{10}(r)) = \emptyset$, it holds that $tr_{10}(r) \in fP_P^E(M)^I$ while $I \not\models H_{hex}(tr_{10}(r))$. Hence, in contradiction to the assumption, it holds that $I \not\models fP_P^E(M)^I$.

(iii) $e2(r) \in I$ and $ne2(r) \in I$: This is similar to case (i), just replace $e1$ by $e2$ and $tr_{10}(r)$ by $tr_{11}(r)$.

(iv) $e2(r) \notin I$ and $ne2(r) \notin I$: This is similar to case (ii), just replace $e1$ by $e2$ and $tr_{10}(r)$ by $tr_{11}(r)$.

Since each case yields a contradiction, it follows that $I$ consistently encodes an explanation candidate.

□

**Lemma 10.** *If $I$ is an answer set for $P_P^E(M)$ and $E = (E_1, E_2)$ is the explanation candidate consistently encoded by $I$, then $fP_P^E(M)^I$ exactly contains*

1. $TR_{16}(M) \cup \ldots \cup TR_{33}(M)$.
2. $\{tr_{12}(r) \mid r \in E_1\} \cup \{tr_{13}(r) \mid r \in br(M) \setminus E_1\} \cup \{tr_{14}(r) \mid r \in E_2\} \cup \{tr_{15}(r) \mid r \in br(M) \setminus E_2\}$.

*Proof.* Let $I$ be an answer set for $P_P^E(M)$ encoding an explanation candidate $E = (E_1, E_2)$.

1. By the constraint rule (34), it holds that $spoil \in I$, thus rules $TR_{30}(M) \cup \ldots \cup TR_{33}(M)$ are in $fP_P^E(M)^I$. Let $tr \in TR_{30}(M) \cup \ldots \cup TR_{33}(M)$, then it holds that $I \models B_{hex}(tr)$, hence it follows that $I \models H_{hex}(tr)$. Therefore, $I \models B_{hex}(tr')$ and $tr' \in fP_P^E(M)^I$, where $tr' \in TR_{16}(M) \cup \ldots \cup TR_{29}(M)$. Specifically, it holds for $tr_{26}(i)$, where $1 \leq i \leq n$, that $I \models B_{hex}(tr_{26}(i))$, because $spoil \in I$ which implies that $f_{\&con\_out'_i}(I, spoil, pres_i, in_i) = 0$.

2. Let $r \in E_1$. Then $e1(r) \in I$ and $ne1(r) \notin I$ since $I$ consistently encodes $E$. Thus, $I \models B_{hex}(tr_{12}(r))$, therefore $tr_{12}(r) \in fP_P^E(M)^I$. Furthermore, $I \not\models B_{hex}(tr_{13}(r))$, hence $tr_{13}(r) \notin fP_P^E(M)^I$.

Let $r \in br(M) \setminus E_1$. Then $e1(r) \notin I$ and $ne1(r) \in I$ since $I$ consistently encodes $E$. Thus, $I \models B_{hex}(tr_{13}(r))$, therefore $tr_{13}(r) \in fP_P^E(M)^I$. Furthermore, $I \not\models B_{hex}(tr_{12}(r))$, hence $tr_{12}(r) \notin fP_P^E(M)^I$.

The remaining cases for $E_2$ are analogous. □

**Definition 17.** *An interpretation $I$ of $P_P^E(M)$ is called* contradiction-free *(regarding $r1$, $nr1$, $r2$, $nr2$, $pres_i$, $abs_i$) if and only if the following conditions hold:*

$$\begin{aligned} r1(r) \in I \text{ iff } nr1(r) \notin I \qquad & \text{for every } r \in br(M) \\ r2(r) \in I \text{ iff } nr2(r) \notin I \qquad & \text{for every } r \in br(M) \\ pres_i(a) \in I \text{ iff } abs_i(a) \notin I \qquad & \text{for every } a \in OUT_i, 1 \leq i \leq n \end{aligned}$$

*We say thta a contradiction-free interpretation $I$* consistently encodes *a belief state $S = (S_1, \dots, S_n)$ and a pair $(R_1, R_2)$ of sets of bridge rules such that: $a \in S_i$ iff $pres_i(a) \in I$, $r \in R_1$ iff $r1(r) \in I$, and $r \in R_2$ iff $r2(r) \in I$.*

Notice that rule (34) and Lemma 10 ensure that no answer set $I$ of $P_P^E(M)$ is contradiction-free, because it holds that $spoil \in I$ and the rules of $TR_{30}(M) \cup \ldots \cup TR_{33}(M)$ ensure the saturation of $I$. The notion, however, is useful for reasoning about (minimal) models of $fP_P^E(M)^I$.

$P_P^E(M)$ guarantees that a contradiction-free interpretation $I$ that encodes a belief state $S$ and a pair $(R_1, R_2)$ of sets of bridge rules also contains a representation of the set of heads of bridge rules applicable under $S$ and $(R_1, R_2)$, as the following lemma shows.

**Lemma 11.** *Let $I$ be a contradiction-free interpretation that encodes the belief state $S = (S_1, \dots, S_n)$ of $M$, and let $(R_1, R_2)$ such that $R_1, R_2 \subseteq br(M)$. If $I$ is a minimal model of $P \subseteq P_P^E(M)$ such that $TR_{17}(M) \cup \ldots \cup TR_{25}(M)$ is a subset of $P$, then $\{b \in IN_i \mid in_i(b) \in I\} = \{h_b(r) \mid r \in app(br_i(M[R_1 \cup cf(R_2)]), S)\}$ for every $1 \leq i \leq n$.*

*Proof.* ($\subseteq$): Let $b \in \{b \in IN_i \mid in_i(b) \in I\}$ and let $\{r_1, \dots, r_k\} = [(i : b)]$ be the set of bridge rules of $M[R_1 \cup cf(R_2)]$ whose head is $(i : b)$. Since $I \models B_{hex}(tr_{25}(i, b))$, it must hold for some rule $r_j$ with $1 \leq j \leq k$ that $r2(r_j) \in I$ or $body(r_j) \in I$.

In the former case it follows that $r_j \in R_2$ and thus $r_j \in app(br_i(M[R_1 \cup cf(R_2)]), S)$, hence $b \in \{h_b(r) \mid r \in app(br_i(M[R_1 \cup cf(R_2)]), S)\}$.

In the latter case, $body(r_j) \in I$ together with rules $tr_{19}(r_j), \dots, tr_{22}(r_j)$ implies that each literal in the body of $r_j$ is satisfied by the belief state $S$. Furthermore, from $I \models tr_{18}(r_j)$ it follows that $r1(r_j) \in I$, hence $r_j \in R_1$. Therefore, $r_j \in app(br_i(M[R_1 \cup cf(R_2)]), S)$, hence $b \in \{h_b(r_j) \mid app(br_i(M[R_1 \cup cf(R_2)]), S)\}$.

($\supseteq$) Let $b \in app(br_i(M[R_1 \cup cf(R_2)]), S)\}$ and let $\{r_1, \dots, r_k\} = [(i : b)]$ be the bridge rules in $M[R_1 \cup cf(R_2)]$ whose head is $(i : b)$. By definition of applicability, it must hold for some $r_j$ with $1 \leq j \leq k$ that either $r_j \in R_2$ or $r_j \in R_1$ and the body of $r_j$ is satisfied wrt. $S$. In the former case $r2(r_j) \in I$ and by $tr_{24}(r_j)$ it must hold that $in_i(b) \in I$, hence $b \in \{b \in IN_i \mid in_i(b) \in I\}$. In the latter case observe that $S \models r_j$ and as $I$ consistently encodes $S$ and $(R_1, R_2)$, it holds that $I \models B_{hex}(tr_{17}(r_j))$. Therefore $in_i(b) \in I$, hence $b \in \{b \in IN_i \mid in_i(b) \in I\}$. □

**Proof of Theorem 6** Recall the concept of a saturated ("spoiled") interpretation. An interpretation is saturated, if it is a superset of $I_{spoil}$, which is defined as follows:

$$\begin{aligned} I_{spoil} = &\{r1(r), nr1(r), r2(r), nr2(r), body(r) \mid r \in br(M)\} \cup \\ &\{in_i(b) \mid r \in br(M) \wedge h_c(r) = i \wedge h_b(r) = b\} \cup \{spoil\} \cup \\ &\bigcup_{a \in OUT_i} \{pres_i(a), abs_i(a)\} \cup \bigcup_{b \in IN_i} \{in_i(b)\}. \end{aligned}$$

*Soundness* ($\Leftarrow$). Let $I$ be an answer set of $P_P^E(M)$. Then by Lemma 9 $I$ consistently encodes an explanation candidate $E = (E_1, E_2)$ where $E_1 = \{r \in br(M) \mid e1(r) \in I\}$ and $E_2 = \{r \in br(M) \mid e2(r) \in I\}$. We show that $E$ is an explanation of $M$.

Since $I$ is an answer set of $P_P^E(M)$, it is a minimal model of $fP_P^E(M)^I$ and by Lemma 10 all rules of $TR_{16}(M) \cup \ldots \cup TR_{33}(M)$ are in $fP_P^E(M)^I$, so $I$ must be a minimal model of those rules. By rule (34) it follows that $spoil \in I$, therefore for each $tr \in TR_{30}(M) \cup \ldots \cup TR_{33}(M)$ it holds that $H_{hex}(tr) \in I$ since $I \models B_{hex}(tr)$. Therefore, $I_{spoil} \subseteq I$.

Towards a contradiction, assume that $E$ is not an explanation. Then, there exists $(R_1, R_2) \in Rg(E)$ such that $M' \not\models \bot$ holds for $M' = M[R_1 \cup cf(R_2)]$, i.e., $M'$ has an equilibrium $S = (S_1, \ldots, S_n)$.

Consider the interpretation $I_{S,(R_1,R_2)}$ corresponding to $S$ and $(R_1, R_2)$, i.e., $I'$ is a contradiction-free interpretation regarding $r1, nr1, r2, nr2, pres_i, abs_i$ that consistently encodes $S$ and $(R_1, R_2)$. Let $I_{S,(R_1,R_2),E} = I_{S,(R_1,R_2)} \cup \{e1(r) \in I\} \cup \{ne1(r) \in I\} \cup \{e2(r) \in I\} \cup \{ne2(r) \in I\}$ be the interpretation consistently encoding $E$, $S$, and $(R_1, R_2)$. Finally, let $I_{app} = \{in_i(b) \mid b \in app(br_i(M'), S)\} \cup \{body(r) \mid r \in R_1 \wedge S \models r\}$ correspond to the set of bridge rule heads and bodies applicable under $S$. Combining them, we obtain an interpretation $I' = I_{S,(R_1,R_2),E} \cup I_{app}$. Note that $I' \subset I$, since $I$ is saturated and both $I$ and $I'$ consistently encode $E$.

As we show in the following, it holds that $I' \models fP_P^E(M)^I$:

- For every $tr \in TR_{10}(M) \cup TR_{11}(M)$ it holds that $I' \models tr$ since $I \models tr$ and $I$ agrees with $I'$ on atoms $e1, ne1, e2$, and $ne2$.
- For every $tr \in TR_{12}(M) \cup \ldots \cup TR_{15}(M)$ it holds that $I' \models tr$ since $(R_1, R_2) \in Rg(E)$ and $I'$ consistently encodes $(R_1, R_2)$.
- For every $tr \in TR_{16}(M)$ it holds that $I' \models tr$ since $I'$ consistently encodes $S$.
- For every $r \in br(M)$ it holds that $I' \models tr_{17}(r)$ since $I_{app} \subseteq I'$ and $I_{app}$ is defined such that $S \mid\!\!\leadsto r$ and $r \in R_1$ implies that $body(r) \in I'$.
- For every $r \in br(M)$ it holds that $I' \models tr_{18}(r)$ since $body(r) \in I'$ implies $r \in R_1$, hence by $I'$ encoding $(R_1, R_2)$ it follows that $r1(r) \in I'$.
- For every $r \in br(M)$ it holds that $I' \models tr_{19}(r), \ldots I' \models tr_{22}(r)$, because $body(r) \in I'$ only if $S \models r$, hence by $I'$ encoding $S$ the following hold: $H_{hex}(tr_{19}(r)) \in I', \ldots, H_{hex}(tr_{22}(r)) \in I'$.
- For every $r \in br(M)$ it holds that $I' \models tr_{23}(r)$, respectively $I' \models tr_{24}(r)$ since $S \models r$ and $r \in R_1$, respectively $r \in R_2$, implies that $r \in app(br_i(M'), S)$, hence $in_i(b) \in I'$ where $i \in ci(M)$ and $h_b(r) = b$.
- For every head $(i : b)$ of a bridge rule it holds that $I' \models tr_{25}(i, b)$, because: if $in_i(b) \in I'$ for some $i \in ci(M)$, then by definition of $I'$ there exists $r \in app(br_i(M'), S)$ such that one of the following holds:
  - $S \models r$ and $r \in R_1$, which implies that $body(r) \in I'$.
  - $r \in R_2$ and therefore $r2(r) \in I'$.
- $I' \models tr_{26}(i)$ holds for all $1 \leq i \leq n$: By definition of $I_{app}$, it holds that $\{b \mid in_i(b) \in I'\} = app(br_i(M'), S)$ and since $I'$ encodes $S$, it also it holds that $\{a \mid pres_i(a)\} = S_i$. By assumption $S$ is

an equilibrium of $M'$, hence $S_i \in \mathbf{ACC}_i(app(br_i(M'), S))$. Therefore, $f_{\&con\_out'_i}(I', pres_i, in_i) = 1$ and $I' \not\models B_{hex}(tr_{26}(i))$.

- For every $tr \in TR_{27}(M) \cup \ldots \cup TR_{29}(M)$ it holds that $I' \models tr$ since $I'$ is conflict-free and $I' \not\models B_{hex}(tr)$.
- For every $tr \in TR_{30}(M) \cup \ldots \cup TR_{33}(M)$ it holds that $I' \models tr$ since $spoil \notin I'$.
- Rule (34): is not in the reduct $fP_P^E(M)^I$, hence it needs not be satisfied by $I'$.

Therefore, all rules of $fP_P^E(M)^I$ are satisfied and it follows that $I'$ is a model of $fP_P^E(M)^I$. Since $I' \subset I$, $I$ is not a minimal model of $fP_P^E(M)^I$, which contradicts that $I$ is an answer set of $P_P^E(M)$. This proves that $E \in E^{\pm}(M)$. $\square$

*Completeness* ($\Rightarrow$). Let $E = (E_1, E_2) \in E^{\pm}(M)$. Then for every $(R_1, R_2) \in Rg(E)$ it holds that $M' \models \bot$ where $M' = M[R_1 \cup cf(R_2)]$, i.e., for every belief state $S = (S_1, \ldots, S_n)$ exists some $1 \leq i \leq n$ such that $S_i \notin \mathbf{ACC}_i(app(br_i(M'), S))$.

We show that $I_E = \{e1(r) \mid r \in E_1\} \cup \{ne1(r) \mid r \in br(M) \setminus E_1\} \cup \{e2(r) \mid r \in E_2\} \cup \{ne2(r) \mid r \in br(M) \setminus E_2\} \cup I_{spoil}$ is an answer set of $P_P^E(M)$.

Since $I_E$ contains respective instances for $e1, ne1, e2$, and $ne2$, $fP_P^E(M)^{I_E}$ contains the following rules: $tr_{12}(r)$ such that $r \in E_1$; $tr_{13}(r)$ such that $r \in br(M) \setminus E_1$; $tr_{14}(r)$ such that $r \in E_2$; and $tr_{15}(r)$ such that $r \in br(M) \setminus E_2$. Furthermore, because $I_E$ contains $I_{spoil}$, $fP_P^E(M)^{I_E}$ contains all rules in $TR_{10}(M) \cup TR_{11}(M) \cup TR_{16}(M) \cup \ldots \cup TR_{33}(M)$. Given that $I_{spoil} \subset I_E$, it is easy to see that $I_E$ is a model of $fP_P^E(M)^{I_E}$. It remains to show that $I_E$ is a $\subseteq$-minimal model of $fP_P^E(M)^{I_E}$.

Assume for contradiction that some $I' \subset I_E$ is a model of $fP_P^E(M)^{I_E}$. Observe that $I_E$ consistently encodes $E$ by definition. Since it must hold that $I' \models tr$ where $tr \in TR_{10}(M) \cup TR_{11}(M)$ and $I' \subset I_E$, it follows that $I'$ also consistently encodes $E$.

Since $fP_P^E(M)^{I_E}$ contains rules $TR_{30}(M) \cup \ldots \cup TR_{33}(M)$ which must be satisfied by $I'$ either $spoil \notin I'$ or all respective heads are in $I'$, which means that $I'$ is saturated. The latter implies that $I' = I_E$, which contradicts the assumption $I' \subset I_E$, it follows that $spoil \notin I'$. This requires that $I' \not\models B_{hex}(tr)$ where $tr \in TR_{26}(M) \cup TR_{27}(M) \cup TR_{28}(M) \cup TR_{29}(M)$.

Since it holds that $I \not\models B_{hex}(tr_{26}(i))$ for all $1 \leq i \leq n$, there exists a contradiction-free guess regarding $r1, nr1, r2, nr2, pres_i, abs_i$ such that $f_{\&con\_out'_i}(I', pres_i, in_i) = 1$. Let $S = (S_1, \ldots, S_n)$ be the belief state consistently encoded by $I'$ and let $(R_1, R_2)$ be the pair of sets of bridge rules consistently encoded by $I'$. It holds that $(R_1, R_2) \in Rg(E)$, because $TR_{12}(M)$ and $TR_{14}(M)$ together with the fact that $I'$ is contradiction-free ensure: $e1(r) \in I'$ implies $r1(r) \in I'$ and $r2(r) \in I'$ implies that $ne2(r) \in I'$. In other words, $R_1 \subseteq E_1$ and $R_2 \subseteq br(M) \setminus E_2$, hence $(R_1, R_2) \in Rg(E)$.

By Lemma 11, $\{b \in IN_i \mid in_i(b) \in I'\} = \{h_b(r) \mid r \in app(br_i(M[R_1 \cup cf(R_2)]), S)\}$ for every $1 \leq i \leq n$, which implies that $S_i \in \mathbf{ACC}_i(\{h_b(r) \mid r \in app(br_i(M[R_1 \cup cf(R_2)]), S)\})$; i.e., $S$ is an equilibrium of $M[R_1 \cup cf(R_2)]$. Since $(R_1, R_2) \in Rg(E)$, this contradicts that $E$ is an explanation of $M$. It follows that no $I' \subset I_E$ is a model of $fP_P^E(M)^{I_E}$. Hence $I_E$ is an answer set of $P_P^E(M)$. $\square$