Proceedings of the Tenth Annual Conference on Advances in Cognitive Systems Submitted 9/2022; published 11/2022

Benchmark based Vitality of Axioms and Preconditions for Datalog Theory Repair

Xue Li XUE.SHIRLEY.LI@ED.AC.UK
Alan Bundy A.BUNDY@ED.AC.UK
Ruiqi Zhu RUIQI.ZHU@ED.AC.UK
School of Informatics, University of Edinburgh, UK

Abstract

In general, axioms a logical theory are not equal in terms of their informational value (IV), and neither are the preconditions in a logical rule. Measuring IV is crucial, particularly in automated repair systems, e.g., the Abduction, Belief Revision and Conceptual Change (ABC) repair system (Li & Bundy, 2022b), because when there are multiple repairs, the one that changes the item of least IV is preferred. However, quantifying IV is challenging. Given a benchmark, we evaluate the IV from the perspective of how much an axiom/precondition supports the benchmark, which is quantitatively defined as the *vitality* of axioms and preconditions. The bigger contribution of an axiom/precondition in supporting the benchmark, the more information it conveys, so the bigger its vitality is. Our evaluation shows that the ABC repair system finds the best-repaired theories with smaller search space by only changing the least vital axioms/preconditions, as shown by our evaluation.

1. Introduction

Automated agents use a representation of their environment (i) to interpret incoming sensory data, (ii) to infer new knowledge from old, (iii) to make plans to achieve their goals, and (iv) to predict the consequences of their actions and those of other agents. These environmental representations are not static. They must change (i) when the environment changes, (ii) when the agent must deal with new kinds of goals, or (iii) when the agent detects that they are erroneous.

Automated theory repair systems have previously been proposed to address representation changes based on the proof of falsehoods (Gärdenfors, 2003; Cox & Pietrzykowski, 1986; Muggleton, 2015). However, there are usually multiple repairs for one fault and a faulty theory can have multiple faults. As a result, these theory repair systems suffer from overproduction, i.e., they may come up with multiple repair solutions and need to determine which one is the best.

We categorise the related work in terms of evaluating the value of an axiom into three groups below. Work in the first group defines the epistemic entrenchment to represent the IV of beliefs (Gärdenfors, 1988) but they only order it without a quantitatively measurement. The second group includes measurements, which only consider inconsistency-like, unwanted consequences while ignoring wanted consequences. The third group emphasises the impact of individual axioms w.r.t. both unwanted consequences and wanted consequences.

1. Epistemic entrenchment (EE) describes the overall IV of a belief (Gärdenfors, 1988). In a logical theory, the more entrenched an axiom, the more valuable it should be, and the less inclined a system is to change it. However, an unaided computer system cannot typically grasp the semantics of theories and therefore cannot judge the information value of each element of that theory. For example, an axiom from a paper published in the best journal in its field is usually considered more entrenched than one published in an unknown workshop. However, a repair system, which does not have the relevant publication knowledge, will not be able to make that inference, which requires background information, commonsense and domain knowledge. To the best of our knowledge, there is no measurement of epistemic entrenchment, but postulates about EE ordering by (Gärdenfors, 1988; Meyer et al., 2000).

2. Some measurements have been proposed to evaluate alternative repair solutions, among which the ones that cause the least information loss are preferred. Some measures consider the size of subsets with specific properties, e.g., the smallest incoherence-preserving subset of TBox in description logic (Schlobach et al., 2003), the probability of a formula measured according to maximal subsets that do not contain any unwanted proofs[1] and the proportion of the language involved in the inconsistency (Hunter et al., 2006), and the responsibility defined to be proportional to the longest proof's length (Meliou et al., 2010). However, these measures only consider ,

3. Nikitina et al. (2012) and Urbonas et al. (2020) measure the impact of an axiom by considering both wanted and unwanted theorems. However, they only take the existence of theorems into account, but not look into the details of the number of proofs for each wanted/unwanted theorems.

In contrast, given a benchmark of both wanted and unwanted theorems, our defined vitality can be measured as a score, which is more feasible in applications. Also, our vitality considers the changes in the number of proofs of wanted/unwanted theorems, which are also significant in order to make repaired theories robust.

In this paper, we define the vitality of both axioms and rule preconditions in a logical theory and demonstrates its use to select those repairs that minimise informational loss. Similarly to (Nikitina et al., 2012), our method reduces the background knowledge into partial observations given as the benchmark: a set of positive examples and a set of negative examples, which makes the computation feasible. Fault numbers describe how flawed the current theory is, while the number of wanted/unwanted proofs represents the potential of the theory being faulty in the future. Thus, unlike (Nikitina et al., 2012), which only considers the fault numbers, the change in the number of proofs is also taken into account in our measure.

Our measurement of vitality helps to select best theories from alternatives, based on their overall vitality scores[2]. To illustrate the performance of this selection, we choose the Abduction, Belief Revision and Conceptual Change (ABC) repair system (Li & Bundy, 2022b; Li et al., 2018; Li,

1. In this paper, a proof is a minimal set that entails a goal.
2. Our code and data are available at GitHub: `https://github.com/XuerLi/Publications/tree/main/ACS2022`.

2021) as the framework of our measure, which has various applications (Tang, 2016; Cai & Bundy, 2022; Bundy et al., 2021; Li & Bundy, 2022a), but suffers from the overproduction of repairs: (i) adding/deleting whole axioms, (ii) adding/deleting preconditions of rules, and/or (iii) changing the theory's signature, e.g., renaming predicates or constants, increasing or decreasing a predicate's arity.

In the rest of paper, we first introduce the background of our approach in §2, followed by our hypothesis in §3, and then define and measure vitality of axioms and preconditions in §4. The above hypothesis is evaluated in §5, followed by our conclusion in §6.

2. Background

We will introduce the background of this paper starting with ABC's main input: Datalog theories, followed by the key definitions and the framework of the ABC repair system.

2.1 Datalog Theories

Datalog is a logic programming language consisting of Horn clauses in which there are no functions except constants. We use this notation to define a subset of first-order logic that we also call *Datalog*. We represent clauses in Kowalski normal form, shown in Definition 2.1 below.

Definition 2.1 (Datalog Formulae)**.**

Let the language *of a Datalog theory $\mathbb{T}$ be a triple $\langle \mathcal{P}, \mathcal{C}, \mathcal{V} \rangle$, where $\mathcal{P}$ are the propositions, C are the constants and V are the variables. We will adopt the convention that variables are written in lower case, and constants and predicates start with a capital letter*[3]*. A proposition is a formula of the form $P(t_1, \ldots, t_n)$, where $t_j \in \mathcal{C} \cup \mathcal{V}$ for $1 \leq j \leq n$, i.e., there are no compound terms. Let $R \in \mathcal{P}$ and $Q_i \in \mathcal{P}$ for $0 \leq i \leq m$ in $\mathbb{T}$. R is called the* head *of the clause and the conjunction of the Q_is forms the* body.

Implication: $(Q_1 \wedge \ldots \wedge Q_m) \implies R$. *These usually represent the rules of $\mathbb{T}$.*

Assertion: $\implies R$. *These usually represent the facts of $\mathbb{T}$.*

Goals: $Q_1 \wedge \ldots \wedge Q_m \implies .$ *These usually arise from the negation of the conjecture to be proved and from subsequent subgoals in a derivation.*

Empty Clause: $\implies .$ *This represents false, which is the target of a refutation-style proof. Deriving it, therefore, represents success in proving a conjecture.*

The Datalog *safety condition* requires that every variable that appears in the head of a clause also appears in the body. Variables in the head but not the body are called *orphans*[4]. There are other Datalog restrictions, but these are to make it behave efficiently as a programming language

3. The opposite of the Prolog convention.
4. Although orphans cannot appear in a well-formed Datalog theory, we define them here because they may be created temporarily during the repair process, so must be identified and then eliminated by subsequent repairs.

and we do not need to adopt them. As we will see, despite these restrictions, Datalog is sufficiently expressive for many practical applications.

A small Datalog theory is given in Example 2.1. The axioms assert that all birds can fly and are feathered; penguins are birds; Tweety and Polly are both birds and Polly can fly.

Example 2.1. *Bird Theory $\mathbb{T}_b$ and its vitality.*

$$
\begin{aligned}
bird(X) &\implies fly(X) && \text{(A1)}\\
bird(X) &\implies feathered(X) && \text{(A2)}\\
penguin(Y) &\implies bird(Y) && \text{(A3)}\\
&\implies penguin(tweety) && \text{(A4)}\\
&\implies bird(polly) && \text{(A5)}\\
&\implies fly(polly) && \text{(A6)}
\end{aligned}
$$

2.2 The ABC Repair System

ABC takes two inputs: a theory $\mathbb{T}$ written in the Datalog language and a *preferred structure* ($\mathbb{PS}$), which refers to the relation constructed by different sets of propositions based on their truth values according to the user, as defined below.

Definition 2.2 (Preferred Structure)**.** *A preferred structure is a pair of structures constructed over the signature of a logical theory ($\mathbb{T}$):*

True Set ($\mathcal{T}(\mathbb{PS})$): *The set of the ground propositions which should be provable by $\mathbb{T}$. These propositions are called the* preferred propositions.

False Set ($\mathcal{F}(\mathbb{PS})$): *The set of the ground propositions which should not be proved by $\mathbb{T}$. These propositions are called the* violative propositions

The preferred structure is the benchmark of the correctness of the input theory.

Figure 1 shows ABC's workflow. The inputs to ABC are a Datalog theory $\mathbb{T}$ and the preferred structure $\mathbb{PS}$ which consists of a pair of sets of ground propositions: those propositions that are observed to be true $\mathcal{T}(\mathbb{PS})$ and those observed to be false $\mathcal{F}(\mathbb{PS})$. The pre-process in C1 reads and rewrites inputs into the internal format for later use. Then in C2, ABC applies selected literal resolution (SL) (Kowalski & Kuehner, 1971) to $\mathbb{T}$ to detect incompatibility and insufficiency faults based on $\mathcal{F}(\mathbb{PS})$ and $\mathcal{T}(\mathbb{PS})$, defined below.

Given a preferred structure $\mathbb{PS}$, a theory $\mathbb{T}$ could have two kinds of faults:

Incompatibility: Predictions that arise from the agent's representation conflict with observations of their environment: $\exists\phi.\ \mathbb{T} \vdash \phi \wedge \phi \in \mathcal{F}(\mathbb{PS})$.

Insufficiency: The agent fails to predict observations of its environment: $\exists\phi.\ \mathbb{T} \nvdash \phi \wedge \phi \in \mathcal{T}(\mathbb{PS})$

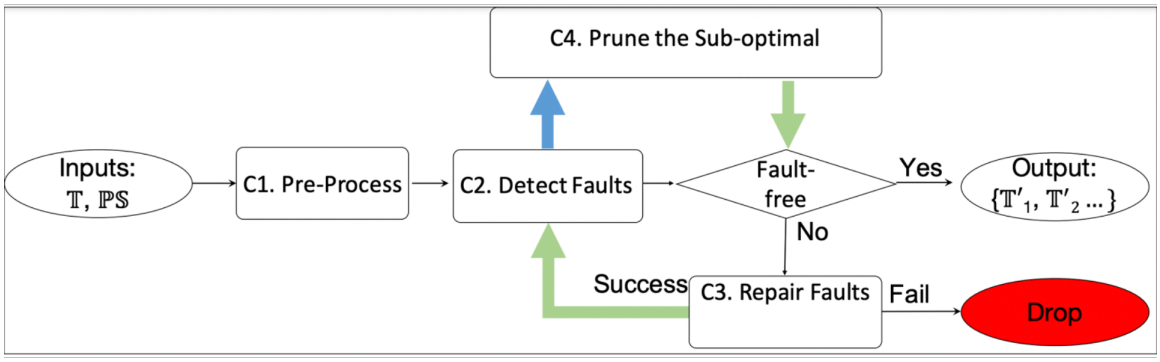


Figure 1. Flowchart of the ABC: green arrows deliver a set of theories one by one to the next process; the blue arrow collects and delivers theories as a set; When a faulty-theory is not repairable, it will be dropped from the repair process.

As ABC uses SL which is not only sound and complete (Gallier, 2003), but also decidable (Pfenning, 2006) for Datalog theories so that proofs can always be detected if there are any.

In C3, repairs are generated to fix detected faults. An insufficiency is repaired by unblocking a proof with additional necessary SL steps, while an incompatibility is repaired by blocking all its proofs, which can be done by breaking one SL step in each of them (Li et al., 2018). ABC repairs faulty theories using eleven *repair operations*. There are five for repairing incompatibilities and six for repairing insufficiencies, defined below.

Definition 2.3 (Repair Operations for Incompatibility)**.** *In the case of incompatibility, the unwanted proof can be blocked by causing any of the SL steps to fail. Suppose the targeted SL step is between a goal, $P(s_1, \ldots, s_n)$, and an axiom, $Body \implies P(t_1, \ldots, t_n)$, where each s_i and t_i pair can be unified. Possible repair operations are as follows:*

Belief Revision 1: *Delete the targeted axiom: $Body \implies P(t_1, \ldots, t_n)$.*

Belief Revision 2: *Add an additional precondition to the body of an earlier rule axiom which will become an unprovable subgoal in the unwanted proof.*

Reformation 3: *Rename P in the targeted axiom to either a new predicate or a different existing predicate P'.*

Reformation 4: *Increase the arity of all occurrences P in the axioms by adding a new argument. Ensure that the new arguments in the targeted occurrence of P, are not unifiable. In Datalog, this can only be ensured if they are unequal constants at the point of unification.*

Reformation 5: *For some i, suppose s_i is C. Since s_i and t_i unify, t_i is either C or a variable. Change t_i to either a new constant or a different existing constant C'.*

Definition 2.4 (Repair Operations for Insufficiency)**.** *In the case of insufficiency, the wanted but failed proof can be unblocked by causing a currently failing SL step to succeed. Suppose the chosen*

SL step is between a goal $P(s_1, \ldots, s_m)$ and an axiom $Body \implies P'(t_1, \ldots, t_n)$, where either $P \neq P'$ or for some i, s_i and t_i cannot be unified. Possible repair operations are:

Abduction 1: *Add the goal $P(s_1, \ldots, s_m)$ as a new assertion and replace variables with constants.*

Abduction 2: *Add a new rule whose head unifies with the goal $P(s_1, \ldots, s_m)$ by analogising an existing rule or formalising a precondition based on a theorem whose arguments overlap with the ones of that goal.*

Abduction 3: *Locate the rule axiom whose precondition created this goal and delete this precondition from the rule.*

Reformation 4: *Replace $P'(t_1, \ldots, t_n)$ in the axiom with $P(s_1, \ldots, s_m)$.*

Reformation 5: *Suppose s_i and t_i are not unifiable. Decrease the arity of all occurrences P' by 1 by deleting its i^{th} argument.*

Reformation 6: *If s_i and t_i are not unifiable, then they are unequal constants, say, C and C'. Either (a) rename all occurrences of C' in the axioms to C or (b) replace the offending occurrence of C' in the targeted axiom by a new variable.*

An example of $\mathbb{PS}$ is given in Example 2.2, where the left side is the original theory $\mathbb{T}_b$. It can be seen that $\mathbb{T}_b$ is incompatible because it proves $fly(tweety)$ from $\mathcal{F}(\mathbb{PS})_b$. One of ABC's repair is given on the right side, where the arity of $bird$ is increased by 1.

Example 2.2. *Bird Theory $\mathbb{T}_b$ on the left and its Repair $\mathbb{T}_r$ on the right.*

$$
\begin{array}{rcr|rcr}
bird(X) & \implies & fly(X) & bird(X, normal) & \implies & fly(X) \\
bird(X) & \implies & feathered(X) & bird(X, Y) & \implies & feathered(X) \\
penguin(Y) & \implies & bird(Y) & penguin(X) & \implies & bird(X, abnormal) \\
 & \implies & penguin(tweety) & & \implies & penguin(tweety) \\
 & \implies & bird(polly) & & \implies & bird(polly, normal) \\
 & \implies & fly(polly) & & \implies & fly(polly)
\end{array}
$$

$$\mathcal{T}(\mathbb{PS})_b = \{penguin(tweety),\ feathered(tweety),\ fly(polly)\}$$
$$\mathcal{F}(\mathbb{PS})_b = \{fly(tweety)\}$$

Usually a faulty theory requires multiple repairs to be fully repaired. Due to the diverse repairs, ABC tends to be over-productive (Li et al., 2018). Thus, only those with the fewest faults are selected as the optimal among alternatives (Li, 2021; Urbonas et al., 2020) in C4. ABC repeats its repair process until there is no fault left.

Beyond the existence of proofs, the change of the proofs' number represents the potential of the theory being faulty in future. However, the sub-optimal pruning does not consider that change so we will define vitality to take it into account in §4. First, however, we will define our hypothesis.

3. Hypothesis

Diverse repairs make ABC suffer from overproduction (Urbonas et al., 2020). So, it is important to know which ones are the best among all produced repaired theories.

Definition 3.1 (Best-Repaired Theory)**.** *A best-repaired theory (BRT) has the following properties.*

1. *the repaired theory satisfies the benchmark of the theory's correctness*[5]*;*
2. *the applied repair operations on a repaired theory are all necessary in terms of fault-repairing to make the theory fully respect its benchmark*[6]*;*
3. *when the input theory is explainable, the repaired theory and its difference from the original theory $\mathbb{T}$ is also intuitively explainable by human.*

The first two properties for a BRT can be evaluated formally. The third stands for human judgement of which repairs are more intuitive than others[7]. For example, $mum(diana, william, birth)$ can be explained as Diana is William's the birth mother, where $birth$ is added as a repair that represents the type of the motherhood, while a repair that changes $capital(uk, london)$ into $capital(uk)$ is not explainable because UK is not a capital so the new axiom does not make sense, unless $capital$ is explained as something counter-intuition. Thus, if we compare the above two repairs then the first is better.

Based on the defined best repaired theories, our hypothesis is as below.

Hypothesis

Selecting repairs based on vitality, will prioritise the best of repaired theories.

In the next section, the vitality will be defined and measured.

4. Vitality Measuring

One Datalog theory is better than another if it has fewer faults relative to its $\mathbb{PS}$. When two theories have the same number of faults, then one is considered better if it has more proofs of true theorems or fewer proofs of false ones. A proof means a subset of the theory which entails the goal with a certain order (Bundy et al., 2005). However, it is convenient, in this paper, to ignore the order and represent a proof as the set of axioms used in it. The relation between a theory and its $\mathbb{PS}$ can be summarised as follows:

5. The benchmark is the preferred structure defined in Definition 2.2.
6. The faults to be repaired are defined in Definition 2.2.
7. Whether a repaired theory is BRT is independent from our measurement of vitality.

- **The number of insufficiencies**: the number of propositions from $\mathcal{T}(\mathbb{PS})$ that are not theorems of the theory: $|\mathcal{IS}(\mathbb{T}, \mathbb{PS})|$, where $\mathcal{IS}(\mathbb{T}, \mathbb{PS}) = \{\phi \in \mathcal{T}(\mathbb{PS}) | \mathbb{T} \nvdash \phi\}$.
- **The number of incompatibilities**: the number of propositions from $\mathcal{F}(\mathbb{PS})$ that are theorems of the theory: $|\mathcal{IC}(\mathbb{T}, \mathbb{PS})|$, where $\mathcal{IC}(\mathbb{T}, \mathbb{PS}) = \{\phi \in \mathcal{F}(\mathbb{PS}) | \mathbb{T} \vdash \phi\}$.

Let $\mathcal{N}(\mathbb{T},\ \alpha)$ be the number of proofs of α in the theory $\mathbb{T}$, shown in equation (1), where $\mathbb{T} \vdash_\pi \phi$ means that π is a proof of ϕ in theory $\mathbb{T}$.

$$\mathcal{N}(\mathbb{T},\ \alpha) = |\{\pi | \mathbb{T} \vdash_\pi \alpha\}| \tag{1}$$

When proposition α comes from $\mathbb{PS}$, $\mathcal{N}(\mathbb{T},\ \alpha)$ is rewritten as either $\mathcal{N}_t(\mathbb{T},\ \alpha)$ or $\mathcal{N}_f(\mathbb{T},\ \alpha)$.

- If $\alpha \in \mathcal{T}(\mathbb{PS})$, then its proof number is written as $\mathcal{N}_t(\mathbb{T},\ \alpha)$, representing **the degree of a sufficiency** w.r.t. a preferred proposition α.
- If $\alpha \in \mathcal{F}(\mathbb{PS})$, its proof number is written as $\mathcal{N}_f(\mathbb{T},\ \alpha)$, representing **The degree of an incompatibility** w.r.t a violative proposition α.

Then we have the following conclusions.

- $\mathcal{N}_t(\mathbb{T},\ \alpha) = 0 \iff \alpha \in \mathcal{IS}(\mathbb{T}, \mathbb{PS})$
- $\mathcal{N}_t(\mathbb{T},\ \alpha) > 0 \iff \alpha \notin \mathcal{IS}(\mathbb{T}, \mathbb{PS})$
- $\mathcal{N}_f(\mathbb{T},\ \alpha) = 0 \iff \alpha \notin \mathcal{IC}(\mathbb{T}, \mathbb{PS})$
- $\mathcal{N}_f(\mathbb{T},\ \alpha) > 0 \iff \alpha \in \mathcal{IC}(\mathbb{T}, \mathbb{PS})$

In general, the main target of repairing a faulty theory is to reduce $|\mathcal{IS}(\mathbb{T}, \mathbb{PS})|$ and $|\mathcal{IC}(\mathbb{T}, \mathbb{PS})|$ to zero. Going further than that, it is good to have greater $\mathcal{N}_t$ and smaller $\mathcal{N}_f$ so that $|\mathcal{IS}(\mathbb{T}, \mathbb{PS})|$ and $|\mathcal{IC}(\mathbb{T}, \mathbb{PS})|$ is less likely to increase when changes are applied to the theory.

Assume that there are two repairs, ν_1 and ν_2, available for a faulty theory. If ν_1 fixes an insufficiency and can reduce the proof number of an incompatibility, e.g., from 4 to 3, while repair ν_2 fixes the same insufficiency and reduces more proofs of that incompatibility, e.g., from 4 to 1, Then ν_2 is a better repair than ν_1.

Therefore, the sub-tasks for repairing a faulty theory are to enhance sufficiency and to weaken incompatibility.

1. **Sufficiency Enhancement**: increase the number of proofs of a preferred proposition, i.e. increase $\mathcal{N}_t(\mathbb{T},\ \alpha)$.
2. **Incompatibility Weakening**: decrease the number of proofs of a violative proposition, i.e. increase $\mathcal{N}_f(\mathbb{T},\ \alpha)$.

Accordingly, the vitality for axioms and preconditions is measured based on the main tasks of repairing insufficiencies and incompatibilities, and the sub-tasks of enhancing sufficiencies and weakening incompatibilities.

4.1 Axiom Vitality

An axiom's vitality will be evaluated according to how much that axiom supports $\mathbb{PS}$: *the more an axiom supports $\mathbb{PS}$, the more entrenched that axiom is in the theory*. The contribution of an axiom to proving a goal is defined.

Definition 4.1 (Axiom Contribution)**.** *The contribution of an axiom (a) to a goal (ϕ) is the fraction of ϕ's proofs π in which a is involved, as given in equation (2).*

$$c(a,\phi) = \begin{cases} \frac{|\{\pi|\mathbb{T}\vdash_\pi \phi \wedge a\in\pi\}|}{|\{\pi|\mathbb{T}\vdash_\pi \phi\}|},\ \mathbb{T}\vdash\phi \\ 0,\ otherwise \end{cases} \tag{2}$$

where $\mathbb{T}\vdash_\pi \phi$ means that π is a proof of ϕ in theory $\mathbb{T}$ and $a\in\pi$ means that a occurs in proof π.

If $c(a,\phi)=1$, it means that axiom a is involved in all the proofs of ϕ. This is a vital feature of an axiom: its deletion or addition changes fault numbers immediately, rather than just weakening incompatibility or enhancing sufficiency.

If an axiom is not in the theory, i.e. $\beta\notin\mathbb{T}$, then $c(\beta,\ \phi)=0$. However, once β has been added into the theory, then $c(\beta,\ \phi)\geq 0$ because the addition of β could complete so unblock some proofs of ϕ.

Theorem 4.1. *If $c(a,\phi)=1$ and $\phi\in\mathcal{T}(\mathbb{PS})$, deleting or adding a to $\mathbb{T}$ will introduce or repair the insufficiency w.r.t. ϕ, respectively.*

Theorem 4.2. *If $c(a,\phi)=1$ and $\phi\in\mathcal{F}(\mathbb{PS})$, deleting or adding a to $\mathbb{T}$ will repair or introduce the incompatibility w.r.t. ϕ, respectively.*

Accordingly, an axiom's vitality ($\mathcal{V}_a$) is defined based on $\mathbb{PS}$ as below, where $\mathcal{V}_{a1}$ emphasises the effects of deleting/adding a on changing fault numbers and $\mathcal{V}_{a2}$ captures its effects on sufficiency enhancement and incompatibility weakening.

Definition 4.2 (Axiom Vitality ($\mathcal{V}_a$))**.** *Based on the preferred structure $\mathbb{PS}$, the vitality of an axiom a from the theory $\mathbb{T}$, or one that is considered adding to $\mathbb{T}$ is written as $\mathcal{V}_a(a,\mathbb{T},\ \mathbb{PS})=\langle\mathcal{V}_{a1},\ \mathcal{V}_{a2}\rangle$, given by equations (3) and (4).*

$$\mathcal{V}_{a1} = |\{\phi|\phi\in\mathcal{T}(\mathbb{PS})\wedge c(a,\phi)=1\}| - |\{\phi|\phi\in\mathcal{F}(\mathbb{PS})\wedge c(a,\phi)=1\}| \tag{3}$$

$$\mathcal{V}_{a2} = \sum_{\substack{\phi\in\mathcal{T}(\mathbb{PS})\\ c(a,\ \phi)<1}} c(a,\ \phi) - \sum_{\substack{\phi\in\mathcal{F}(\mathbb{PS})\\ c(a,\ \phi)<1}} c(a,\ \phi) \tag{4}$$

The changes in the fault number $\mathcal{V}_{a1}$ are seen as more important than changes in the number of proofs $\mathcal{V}_{a2}$, so that $\mathcal{V}_a$s can be compared based on a lexicographic order $\succ_l$.

Definition 4.3 (Vitality Comparison $\succ_l$)**.** *Let $\mathcal{V}_a(\alpha,\ \mathbb{T},\ \mathbb{PS})=\langle\mathcal{V}_{a1},\ \mathcal{V}_{a2}\rangle$ and $\mathcal{V}_a(\beta,\ \mathbb{T},\ \mathbb{PS})=\langle\mathcal{V}'_{a1},\ \mathcal{V}'_{a2}\rangle$, then α is more entrenched than β , denoted as $\alpha\succ_l\beta$, when*

$$(\mathcal{V}_{a1}>\mathcal{V}'_{a1})\vee(\mathcal{V}_{a1}=\mathcal{V}'_{a1}\wedge\mathcal{V}_{a2}>\mathcal{V}'_{a2}) \tag{5}$$

i.e., compared with deleting β, deleting α leaves more faults, or it leaves the same number of faults but retains more unwanted proofs or breaks more wanted proofs.

When there are multiple solutions to deleting or adding axioms, the one maximising the sum of all axioms' vitality will be chosen. To illustrate this, consider the faulty theory in Example 4.1, where the proofs of the first two propositions in $\mathcal{T}(\mathbb{PS})_b$ are constructed by (A4) and $(A4, A3, A2)$ respectively. And there are two proofs of the third proposition: (A6) and $(A5, A1)$. As for the proposition $fly(tweety)$ in $\mathcal{F}(\mathbb{PS})_b$, its only proof uses $(A4, A3, A1)$. According to Definition 4.2, the vitality of each axiom is calculated.

Example 4.1. *Bird Theory $\mathbb{T}_b$ and its vitality.*

$$
\begin{aligned}
bird(X) &\implies fly(X) && \text{(A1)}\\
bird(X) &\implies feathered(X) && \text{(A2)}\\
penguin(Y) &\implies bird(Y) && \text{(A3)}\\
&\implies penguin(tweety) && \text{(A4)}\\
&\implies bird(polly) && \text{(A5)}\\
&\implies fly(polly) && \text{(A6)}
\end{aligned}
$$

$$\mathcal{T}(\mathbb{PS})_b = \{penguin(tweety), feathered(tweety), fly(polly)\};$$
$$\mathcal{F}(\mathbb{PS})_b = \{fly(tweety)\}.$$

$$\mathcal{V}_a(A1,\ \mathbb{T}_b,\ \mathbb{PS}_b) = \langle -1, 0.5\rangle;\ \ \mathcal{V}_a(A2,\ \mathbb{T}_b,\ \mathbb{PS}_b) = \langle 1, 0\rangle;$$
$$\mathcal{V}_a(A3,\ \mathbb{T}_b,\ \mathbb{PS}_b) = \langle 0, 0\rangle;\qquad \mathcal{V}_a(A4,\ \mathbb{T}_b,\ \mathbb{PS}_b) = \langle 1, 0\rangle;$$
$$\mathcal{V}_a(A5,\ \mathbb{T}_b,\ \mathbb{PS}_b) = \langle 0, 0.5\rangle;\ \ \mathcal{V}_a(A6,\ \mathbb{T}_b,\ \mathbb{PS}_b) = \langle 0, 0.5\rangle$$

So, $(A4)$ and $(A2)$ are the most entrenched, $(A1)$ is least. Between them, $(A5)$ and $(A6)$ are equally entrenched more than $A3$. As a result, $(A1)$ in Example 4.1 will be chosen to be changed when repairing a fault based on $\mathbb{PS}_b$. This example illustrates the choice made w.r.t. deleting an axiom. Similarly, when it comes to adding new axioms, the repair containing the most entrenched new axiom(s) will be selected, as it maximises the overall axiom vitality of the theory.

4.2 Precondition Vitality

In this section, we extend our analysis of vitality to preconditions in a rule, relative to a $\mathbb{PS}$. When a precondition is added to a rule axiom in a theory, some of the original theorems may become unprovable. We define this theorem difference as that precondition's *impact*.

Definition 4.4 (Precondition Impact ($\mathcal{PI}$))**.** *In Datalog theory $\mathbb{T}$, if the rule axiom R's preconditions are $p_i(\vec{t_i})$s, $1 \leq i \leq n$, then the impact of precondition $p_i(\vec{t_i})$ of R is the difference in $\mathbb{T}$'s theorems caused by it.*

$$\mathcal{PI}(p_i(\vec{t_i})) = \{\alpha |\ \alpha \in \mathcal{C}(\mathbb{T}'),\ \alpha \notin \mathcal{C}(\mathbb{T})\} \tag{6}$$

where $R \in \mathbb{T},\ R' = R \overset{*}{-} p_i(\vec{t_i})$ *and* $\mathbb{T}' = (\mathbb{T} \dot{-} R) \dot{+} R'$*; where function* $\overset{*}{-}$ *removes a precondition from a rule; where functions* $\dot{-},\ \dot{+}$ *remove or add one axiom to a set of axioms respectively and* $\mathcal{C}$ *returns all of the theorems of a theory.*

The vitality of a precondition $p(\vec{t})$ is measured based on how its impact overlaps $\mathbb{PS}$ in Definition 4.5, where $\mathcal{V}_{p1}$ asserts that the more insufficiencies are caused by the inclusion of $p(\vec{t})$, the less entrenched $p(\vec{t})$ should be, and the more incompatibilities are caused by the absence of $p(\vec{t})$, the more entrenched $p(\vec{t})$ should be. In addition, the more sufficiencies are enhanced by the absence of $p(\vec{t})$, the less entrenched it should be, and the more incompatibilities are weakened by the inclusion of $p(\vec{t})$, the more entrenched it should be.

Definition 4.5 (Vitality of a Precondition ($\mathcal{V}_p$))**.** *The vitality of a precondition is decided by its impact on* $\mathbb{PS}$*, written as* $\mathcal{V}_p(p(\vec{t}),\ R,\ \mathbb{T},\ \mathbb{PS}) = \langle \mathcal{V}_{p1},\ \mathcal{V}_{p2} \rangle$*, given by equations (7) and (8), where* $\mathbb{T}' = \mathbb{T} \dot{-} R \dot{+} R'$*,* $R' = R \overset{*}{-} p(\vec{t})$*, i.e. the rule* R *in* $\mathbb{T}$ *contains the precondition* $p(\vec{t})$*, but this precondition is removed in* $\mathbb{T}'$*.*

$$\mathcal{V}_{p1} = (\mathcal{IS}(\mathbb{T}', \mathbb{PS}) - \mathcal{IS}(\mathbb{T}, \mathbb{PS})) + (\mathcal{IC}(\mathbb{T}', \mathbb{PS}) - \mathcal{IC}(\mathbb{T}, \mathbb{PS})) \tag{7}$$

$$\mathcal{E}_{p2} = \sum_{\substack{\phi \in \mathcal{T}(\mathbb{PS}) \\ T \vdash \phi}} (1 - \frac{|\{\pi | \mathbb{T}' \vdash_\pi \phi\}|}{|\{\pi | \mathbb{T} \vdash_\pi \phi\}|}) + \sum_{\substack{\phi \in \mathcal{F}(\mathbb{PS}) \\ T \vdash \phi}} (\frac{|\{\pi | \mathbb{T}' \vdash_\pi \phi\}|}{|\{\pi | \mathbb{T} \vdash_\pi \phi\}|} - 1) \tag{8}$$

When there are multiple solutions for precondition changes to one fault, the one maximising overall $\mathcal{V}_P$ for that rule is the highest priority for change.

The vitality comparison function $\succ_l$ given by Definition 4.3 can be applied to a pair of axioms, or preconditions or an axiom and a precondition. For example, if $\mathcal{V}_a(a,\ \mathbb{T},\ \mathbb{PS}) = \langle \mathcal{V}_{a1},\ \mathcal{V}_{a2} \rangle$, $\mathcal{V}_p(p(\vec{t}),\ R,\ \mathbb{T},\ \mathbb{PS}) = \langle \mathcal{V}_{p1},\ \mathcal{V}_{p2} \rangle$, then $a \succ_l p(\vec{t})$ when:

$$(\mathcal{V}_{a1} > \mathcal{V}_{p1}) \vee (\mathcal{V}_{a1} = \mathcal{V}_{p1} \wedge \mathcal{V}_{a2} > \mathcal{V}_{p2}) \tag{9}$$

In our measure, $\mathcal{V}_{a2}$ and $\mathcal{V}_{p2}$ represent the repair's impact on the proof number of sufficiencies and incompatibilities. Among repairs which retain the same number of faults, $\mathcal{V}_{a2}$ and $\mathcal{V}_{p2}$ prioritise more robust repairs: the one that results in more sufficiency proofs will be less likely to become insufficient and one in fewer incompatibility proofs will be more likely to repair easily.

5. Evaluation

Following our hypothesis, and our standard of Best Repairs (BR) in §1, our evaluation will focus on whether changing the least entrenched items leads to the BRT over non-BRT alternatives. We determine this by *comparing the vitality of changed items*, in BRT, and non-BRT.

Recall the definition of BRT given in Definition 3.1, the evaluation criteria are defined as the *Silver Standard* (SS) and the *Gold Standard* (GS). The difference between GS and SS is that a

Theory Name (Reference)	Size	#FN	#VB	#MSCR	#NonOpt	LE
Families (Bundy & Mitrovic, 2016)	4-2-0	1-0	2-4	2-5	NES	Y
Tweety (Strasser & Antonelli, 2019)	5-3-1	0-1	1-1	1-1	NES	Y
Researcher (Rodler & Eichholzer, 2019)	4-1-1	0-1	2-8	2-8	2-8	Y
Super Penguin (Gómez et al., 2010)	7-1-1	0-1	2-6	2-6	NES	Y
Buy Stock (Gómez et al., 2010)	9-1-0	0-1	2-8	2-8	NES	Y
Working Student (Gómez et al., 2010)	6-2-0	1-0	1-6	1-6	NES	Y
Missing Parent (Muggleton, 2017)	10-4-5	2-0	1-3	1-8	NES	Y
Parent (Muggleton, 2017)	6-3-3	3-0	1-1	1-4	NES	Y
Capital City (Bundy & Mitrovic, 2016)	5-0-4	0-2	2-12	2-16	2-16	Y
Married Woman (Gómez et al., 2010)	5-1-1	1-1	1-2	1-3	NES	Y

Table 1. Experimental results showing the performance of axiom/precondition vitality in ranking ABC's non-signature repairs.

repaired theory at GS level is human understandable and explainable while one at SS level might not make sense to humans.

Definition 5.1 (Silver Standard and Gold Standard)**.** *The repaired theories that have Properties 1 and 2 in Definition 3.1 are in Silver Standard and the ones have all three properties are in Gold Standard.*

Whether a repair is at GS/SS is checked manually. The tested theories and the chosen of the preferred structure are from the related literature, whose citation is attached to the tested theory names. To make the evaluation as unbiased as possible, our formalisations follow the original source as exactly as feasible, and our adaptations preserve commonsense meanings from the original.

Test results are given in Table 1, whose columns depict the following statistics:

Theory Name: The names of the 10 selected faulty theories, followed by their sources. Adaptation may be involved, e.g., to formalise $\mathbb{PS}$ and translate the original theory into Datalog.

Size: A figure given in the form of X-Y-Z, which are the number of axioms in the theory; the numbers of propositions in $\mathcal{T}(\mathbb{PS})$ and $\mathcal{F}(\mathbb{PS})$, respectively.

#FN: A figure given in the form of X-Y, which are the number of insufficiencies and incompatibilities, respectively.

#VB: A figure given in the form of X-Y, where X is the number of BRTs and Y is the number of all selected repairs **based on vitality**. Thus, this column shows the outputted theories at GS against the ones at the SS.

#MSCR ABC's performance when it selects the optimal theories based on the maximal Set of commutative repairs (MSCR) rather than vitality. Figures are written in the same format as in column #VB.

#NonOpt: ABC's performance without the selection of optimal theories. Figures are written in the same format with #VB and **NES** means never-ending search, i.e., it does not terminate and reaches the default stack limit of SWI Prolog (Wielemaker et al., 2012), which is the program language in which ABC is implemented (Li, 2021).

LE Whether, for each BRT, all the changes are to the least entrenched items.

According to our hypothesis given in §3, a BRT from each of our tests should only repair the least entrenched items. Shown by the last column in Table 1, all of the 10 theories pass the *LE* test, which represents the evaluated vitality prioritise repairs on the least entrenched axioms or preconditions. Thus, it supports our hypothesis. However, there remains the possibility of exceptions, due to the inherently very limited domain and commonsense knowledge formalised in each theory and $\mathbb{PS}$. In addition, ABC provides users with optional heuristics to guide the repair generation, which helps to avoid bad repairs that produce theories at SS and prune ones at GS. In our evaluation, the set of heuristics for individual input theory is kept the same to avoid bias.

Comparison between #VB and #NonVB also shows that selecting repairs based on vitality reduces the search space dramatically. It helps ABC skip bad search branches (BSBs) that are too long or non-terminated, shown as 'NonES' cases. BSBs are caused by the interaction of faults, e.g., a repair can introduce new faults when fixing an old fault.

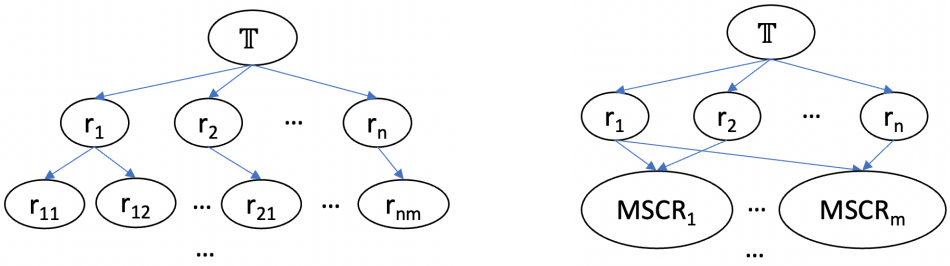
ABC's previous best performance was based on the combination of the optimal selection and applying MSCRs, of which the figure on our tested theories is given in column #MSCR. The optimal selection only keeps the repaired theories with the fewest remaining faults (Urbonas et al., 2020) and MSCR method combines commutative repairs that can be applied together to further reduce the search space of BRTs, which is shown in Figure 2. Instead of considering the remaining faults, our selection using vitality compares the overall vitality scores of repaired theories after applying MSCRs. By comparing figures in the column of #VB and #MSCR in Table 1, it can be seen that the selection of repaired theories based on vitality has better or at least the same performance as the one based on MSCR. When the number of proofs of faults is greater or have a more complex relationship with each other, we would expect a significant advance of vitality.

6. Conclusion

This paper defined the vitality for axioms and preconditions, which are measurable based on the remaining fault number and the number of wanted/unwanted proofs based on a given benchmark. They enable automated repair systems to choose between rival repairs. As an example, the ABC system is discussed. By following the guidance of the defined vitality, ABC only selects the repaired theories where the least vital elements are changed. As a result, these chosen theories not only retain the least faults but also are less likely to be faulty and easier to be repaired in future.

Our measure is based on a limited set of benchmark knowledge: two provided sets, of true and false observations. Thus, it can be applied to any decidable system with positive and negative examples. The future work includes:

1. analysing how sensitive our measure is to the precise way a theory is formulated and whether that reduces the representation power of the vitality;



(a) The original naive search space. (b) The reduced search space based on MSCRs.

The length of each search branch can be different. By applying all repairs in one search branch, that branch terminates with a fault-free theory or with failure, if no repair is available to fix a detected fault.

Figure 2. ABC's Search Space for BRTs.

2. a more sophisticated evaluation, ideally w.r.t. open source data, e.g., Wikidata (Vrandečić & Krötzsch, 2014);

3. a combination mechanism of our vitality of axioms/preconditions and the signature entrenchment (Li et al., 2021), which are independent currently;

4. exploring applications including using the defined vitality as one feature for a more sophisticated measure for evaluating the value of axioms or preconditions. For example, vitality can be extended to also consider how important one theorem in the preferred structure is; or used as a feature together with other features that are representative in terms of evaluating IV, e.g., the trustworthy level of the source of an axiom.

Though we have focused on vitality implemented in the ABC system, we present our approach as a solution for repair selection applicable to other automated theory repair systems. Overall, we believe these algorithms could even be of wider utility beyond repairing faulty representations. For example, the preferred structure is comparable to the sets of positive and negative examples used to guide machine learning (Muggleton, 2017). In that case, a single set of examples might be pressed into double services, refining machine learning classification algorithms while also working to modify their representations, through a repair algorithm such as ABC.

ACKNOWLEDGMENTS

The authors would like to thank Huawei for supporting the authors under grant CIENG4721/LSC. We also acknowledge the support of UKRI grant EP/V026607/1, ELIAI (The Edinburgh Laboratory for Integrated Artificial Intelligence) and EPSRC grant no EP/W002876/1.

References

Bundy, A., Jamnik, M., & Fugard, A. (2005). What is a proof? *Phil. Trans. R. Soc A*, *363*, 2377–2392.

Bundy, A., & Mitrovic, B. (2016). *Reformation: A domain-independent algorithm for theory repair*. Technical report, University of Edinburgh.

Bundy, A., Philalithis, E., & Li, X. (2021). Modelling repairs to virtual bargaining via representational change. *Human-Like Machine Intelligence* (pp. 68–89). Oxford University Press.

Cai, C.-H., & Bundy, A. (2022). Repairing numerical equations in analogically blended theories using reformation. *Proceedings of HLC 2022*. CEUR.

Cox, P. T., & Pietrzykowski, T. (1986). Causes for events: their computation and applications. *International Conference on Automated Deduction* (pp. 608–621). Springer.

Gallier, J. (2003). *Sld-resolution and logic programming*. Chapter 9 of Logic for Computer Science: Foundations of Automatic Theorem Proving. Originally published by Wiley, 1986.

Gärdenfors, P. (1988). *Knowledge in flux: Modeling the dynamics of epistemic states.*. The MIT press.

Gärdenfors, P. (2003). *Belief revision*, volume 29. Cambridge University Press.

Gómez, S. A., Chesñevar, C. I., & Simari, G. R. (2010). Reasoning with inconsistent ontologies through argumentation. *Applied Artificial Intelligence*, *24*, 102–148. From `https://doi.org/10.1080/08839510903448692`.

Hunter, A., Konieczny, S., et al. (2006). Shapley inconsistency values. *KR*, *6*, 249–259.

Kowalski, R. A., & Kuehner, D. (1971). Linear resolution with selection function. *Artificial Intelligence*, *2*, 227–60.

Li, X. (2021). *Automating the repair of faulty logical theories*. Doctoral dissertation, School of Informatics, University of Edinburgh.

Li, X., & Bundy, A. (2022a). ABC repair system in root cause analysis by adding missing information. *The 8th International Online & Onsite Conference on Machine Learning, Optimization, and Data Science, special session of AI for Network/Cloud Management*.

Li, X., & Bundy, A. (2022b). An overview of the abc repair system for datalog-like theories. *International Workshop on Human-Like Computing (HLC 2022)* (p. 30).

Li, X., Bundy, A., & Philalithis, E. (2021). Signature entrenchment and conceptual changes in automated theory repair. *The Ninth Annual Conference on Advances in Cognitive Systems*.

Li, X., Bundy, A., & Smaill, A. (2018). ABC repair system for Datalog-like theories. *KEOD* (pp. 333–340).

Meliou, A., Gatterbauer, W., Moore, K. F., & Suciu, D. (2010). The complexity of causality and responsibility for query answers and non-answers. *arXiv preprint arXiv:1009.2021*.

Meyer, T. A., Labuschagne, W. A., & Heidema, J. (2000). Refined epistemic entrenchment. *Journal of Logic, Language and Information*, *9*, 237–259.

Muggleton, S. H. (2015). Learning efficient logical robot strategies involving composable objects. *Proceedings of the 24th International Joint Conference Artificial Intelligence (IJCAI 2015)* (pp. 3423–3429).

Muggleton, S. H. (2017). Meta-interpretive learning: achievements and challenges. *International Joint Conference on Rules and Reasoning* (pp. 1–6). Springer.

Nikitina, N., Rudolph, S., & Glimm, B. (2012). Interactive ontology revision. *Journal of web semantics*, *12*, 118–130.

Pfenning, F. (2006). *Datalog*. Lecture 26. 15-819K: Logic Programming. From `https://www.cs.cmu.edu/~fp/courses/lp/lectures/26-datalog.pdf`.

Rodler, P., & Eichholzer, M. (2019). On the usefulness of different expert question types for fault localization in ontologies. *Advances and Trends in Artificial Intelligence. From Theory to Practice* (pp. 360–375). Cham: Springer International Publishing.

Schlobach, S., Cornet, R., et al. (2003). Non-standard reasoning services for the debugging of description logic terminologies. *Ijcai* (pp. 355–362). Citeseer.

Strasser, C., & Antonelli, G. A. (2019). Non-monotonic logic. In E. N. Zalta (Ed.), *The stanford encyclopedia of philosophy*. Metaphysics Research Lab, Stanford University, summer 2019 edition.

Tang, J. W. Q. (2016). *Arithmetic errors revisited: Diagnosis and remediation of erroneous arithmetic performance as repair of faulty representations*. Msc thesis, School of Informatics, University of Edinburgh.

Urbonas, M., Bundy, A., Casanova, J., & Li, X. (2020). The use of max-sat for optimal choice of automated theory repairs. *Artificial Intelligence XXXVII* (pp. 49–63). Cham: Springer International Publishing.

Vrandečić, D., & Krötzsch, M. (2014). Wikidata: a free collaborative knowledgebase. *Communications of the ACM*, *57*, 78–85.

Wielemaker, J., Schrijvers, T., Triska, M., & Lager, T. (2012). Swi-prolog. *Theory and Practice of Logic Programming*, *12*, 67–96.