Automated Reasoning for the Dialetheic Logic RM3

Francis Jeffry Pelletier
Department Philosophy
University of Alberta

Geoff Sutcliffe
Department Computer Science
University of Miami

Allen P. Hazen
Department Philosophy
University of Alberta

Abstract

This paper describes a system for automated reasoning in the dialetheic logic RM3. A dialetheic logic allows formulae to be true, or false, or (differently from classical logic) both true and false, and the connectives are interpreted in terms of these three truth values. Consequently some inference rules of classical logic are invalid in RM3, and some theorems of classical logic are not theorems of RM3. An automated theorem prover for first-order RM3 has been developed, based on translations of RM3 formulae to classical first-order logic, and use of existing first-order reasoning systems to reason over the translated formulae. Examples and results are provided to highlight the differences between reasoning in RM3 and classical logic.

Introduction

A logic is *paraconsistent* if it does not allow the derivation of all formulae from the presence of a contradiction. That is, a logic is *paraconsistent* if it does not contain $\phi, \neg\phi \vdash \psi$ as a valid rule of inference (for arbitrary ϕ and ψ). The rationale for paraconsistency is easy to grasp: if logic is to be the correct guide to reasoning, then it seems wrong to say that an agent (human, programmed, whatever) should infer every sentence when there is a contradiction in the input, especially if the agent doesn't realize there is a contradiction. From a somewhat different point of view, if logic is viewed not a normative theory but rather an empirical account of how people reason, it is clear that no one is willing to infer all conclusions from believed input, even when that input is contradictory.

Paraconsistent logics are also used in *dialetheic theories* – theories that allow contradictions to be true (as well as false). As such, statements can be true, or false, or (differently from classical logic) both true and false (simultaneously, in the same respect, at the same time, etc., etc.). The examples that are usually cited in the literature of such dialetheias concern semantic paradoxes, such as the Liar Paradox, where the sentence "This sentence is false." is false if it is true, and is true if it is false. Examples from set theory are the Paradox of Well-Founded sets, the Paradox of a Universal Set, Russell's Paradox, and Richard's Paradox (see (Cantini 2014) for a historical overview of these). There are also other categories of examples that have been discussed in philosophy over the ages: Can God make a stone so heavy that He can't lift it? Perhaps the answer is, "Well, yes and no." There are cases of vague predicates such as *green* and *religion*, e.g., an object can be both green and not green, and a belief system can be both a religion and not a religion. There are legal systems that are inconsistent, in which an action is both legal and illegal. In the physical world, there is a point when a person is walking through a doorway at which the person is both in and not in the room. And so on.

This paper describes a system for automated reasoning in the dialetheic logic RM3. The system is based on translations of RM3 formulae to classical first-order logic (FOL), and subsequent use of existing first-order reasoning systems. Examples and results are provided to highlight the differences between reasoning in RM3 and classical logic.

The Logic LP

The most famous – and persistent – advocate of dialetheism is Graham Priest, who not only examines the status of these informal paradoxes, but also rather long ago developed *The Logic of Paradox*, LP, to accommodate this view (Priest 2006). Although proponents of dialetheic logics like LP insist that there are only two truth values – true and false, and that a sentence can just *be* true, or false, or both, LP is normally presented as a 3-valued logic with three truth values: $\mathbf{T}_3$, $\mathbf{B}_3$, and $\mathbf{F}_3$ (the subscript "3" is used to differentiate them from the two classical truth values, $\mathbf{T}_2$ and $\mathbf{F}_2$). Sentences can be only true ($\mathbf{T}_3$), only false ($\mathbf{F}_3$), or both ($\mathbf{B}_3$). The semantics of the LP connectives extend those of classical logic to provide values for $\mathbf{B}_3$, as given in Table 1, e.g., $\mathbf{B}_3 \rightarrow \mathbf{F}_3$ is $\mathbf{B}_3$. The *designated values* of LP are both $\mathbf{T}_3$ and $\mathbf{B}_3$, so that an interpretation is a model of a formula if the formula is either $\mathbf{T}_3$ or $\mathbf{B}_3$. (If the designated value were just $\mathbf{T}_3$ then the system would be simply Kleene's "strong" 3-valued logic (Kleene 1952, §64).) As usual, a conjecture ψ is a logical consequence of input axioms ϕ if every model of ϕ is a model of ψ, written $\phi \models \psi$. If ϕ is the empty set then ψ is a theorem of the logic, written $\models \psi$.

There are features of LP that have made even dialetheists uneasy with this logic. The main problem is that LP's conditional does not support the rules of inference *modus ponens* and *modus tollendo ponens*. That is, the inferences $\phi, (\phi \rightarrow \psi) \models \psi$ and $\neg\phi, (\phi \vee \psi) \models \psi$ are invalid, as can

be seen by assigning ϕ the value $\mathbf{B}_3$ and ψ the value $\mathbf{F}_3$. This is despite the fact that the corresponding implications $(\phi \wedge (\phi \rightarrow \psi)) \rightarrow \psi$ and $(\neg\phi \wedge (\phi \vee \psi)) \rightarrow \psi$ are theorems of LP. In fact, the theorems of LP are identical to those of classical logic – LP does nothing more than divide up the classical notion of truth into two parts: the "true-only" and the "true-and-also-false". As both types of truth are designated values in the logic, the theorems of the two are the same, and hence LP is not different from classical logic so far as logical truth goes.

In (2008, pp.125-127) Priest lists a number of formulae and rules of inference involving the LP conditional, usually with the idea that these are bad or logico-philosophically indefensible features of classical logic, and states whether or not they are theorems or valid rules of inference in four different many-valued logics, including LP. Even though his heart is with LP, he admits (p.127) that "about the best of the bunch is RM3", with respect to how they treat the conditional. And it is to that logic this paper now turns.

The Logic RM3

RM3 is a paraconsistent logic, and hence a dialetheia-tolerating logic. It differs from LP only in its treatment of the (bi-)conditional, which leads to various further consequences in the overall shape of the logic. The semantics of the RM3 connectives are given in Table 1, with the different values for LP's (bi-)conditional in parentheses. Ordering the truth values $\mathbf{T}_3 > \mathbf{B}_3 > \mathbf{F}_3$, the conditional in RM3 enforces the rule that if the transition from the antecedent to the consequent lowers the truth-value, then the conditional is $\mathbf{F}_3$. This has the effect of making the rule of inference *modus ponens* valid. (The invalidating case for LP, when ϕ is $\mathbf{B}_3$ and ψ is $\mathbf{F}_3$, no longer makes the premises $\mathbf{B}_3$). On the other hand, the rule *modus tollendo ponens* remains invalid: when ϕ is $\mathbf{B}_3$ and ψ is $\mathbf{F}_3$ $\neg\phi, (\phi \vee \psi) \vDash \psi$ has both premises $\mathbf{B}_3$ and the conclusion $\mathbf{F}_3$. In FOL (and LP) the two rules of *modus ponens* and *modus tollendo ponens* stand together, because $\phi \rightarrow \psi$ and $\neg\phi \vee \psi$ are equivalent. In RM3 they are not equivalent, and as a result, while all theorems of RM3 are also theorems of FOL, some theorems of FOL are not theorems of RM3.

ϕ	ψ	$\neg\phi$	$(\phi \wedge \psi)$	$(\phi \vee \psi)$	$(\phi \rightarrow \psi)$	$(\phi \leftrightarrow \psi)$
$\mathbf{T}_3$	$\mathbf{T}_3$	$\mathbf{F}_3$	$\mathbf{T}_3$	$\mathbf{T}_3$	$\mathbf{T}_3$	$\mathbf{T}_3$
$\mathbf{T}_3$	$\mathbf{B}_3$		$\mathbf{B}_3$	$\mathbf{T}_3$	$\mathbf{F}_3$ ($\mathbf{B}_3$)	$\mathbf{F}_3$ ($\mathbf{B}_3$)
$\mathbf{T}_3$	$\mathbf{F}_3$		$\mathbf{F}_3$	$\mathbf{T}_3$	$\mathbf{F}_3$	$\mathbf{F}_3$
$\mathbf{B}_3$	$\mathbf{T}_3$	$\mathbf{B}_3$	$\mathbf{B}_3$	$\mathbf{T}_3$	$\mathbf{T}_3$	$\mathbf{F}_3$ ($\mathbf{B}_3$)
$\mathbf{B}_3$	$\mathbf{B}_3$		$\mathbf{B}_3$	$\mathbf{B}_3$	$\mathbf{B}_3$	$\mathbf{B}_3$
$\mathbf{B}_3$	$\mathbf{F}_3$		$\mathbf{F}_3$	$\mathbf{B}_3$	$\mathbf{F}_3$ ($\mathbf{B}_3$)	$\mathbf{F}_3$ ($\mathbf{B}_3$)
$\mathbf{F}_3$	$\mathbf{T}_3$	$\mathbf{T}_3$	$\mathbf{F}_3$	$\mathbf{T}_3$	$\mathbf{T}_3$	$\mathbf{F}_3$
$\mathbf{F}_3$	$\mathbf{B}_3$		$\mathbf{F}_3$	$\mathbf{B}_3$	$\mathbf{T}_3$	$\mathbf{F}_3$ ($\mathbf{B}_3$)
$\mathbf{F}_3$	$\mathbf{F}_3$		$\mathbf{F}_3$	$\mathbf{F}_3$	$\mathbf{T}_3$	$\mathbf{T}_3$

Table 1: Interpretation of connectives in RM3 (and LP)

First-order RM3 adds predicates, functions, and quantified variables to propositional RM3. Universally quantified formulae are interpreted as the conjunction of the instances of the formula (and thus have the least of the truth values of the conjuncts), and existentially quantified formulae are interpreted as the disjunction of the instances of the formula (and thus have the greatest of the truth values of the disjuncts). An interpretation in RM3 has the same structure as in FOL: a domain $\mathcal{D}$, a function for each function symbol mapping tuples of domain elements to domain elements, and for each predicate an assignment of tuples of domain elements into two sets: the extension – the tuples for which the predicate is interpreted as true, and the anti-extension – the tuples for which it is interpreted as false. As in FOL the extension and anti-extension exhaust the domain. However, in contrast to FOL, the extension and anti-extension are not necessarily disjoint: a tuple of domain elements $<d_1, \ldots, d_n>$ might be in both the extension and the anti-extension of a predicate. The RM3 interpretation of a ground atom $\mathcal{P}(t_1 \ldots t_n)$, with the terms t_i interpreted as the domain elements d_i, is $\mathbf{T}_3$ if $<d_1, \ldots, d_n>$ is in only the extension of $\mathcal{P}$, $\mathbf{F}_3$ if the tuple is in only the anti-extension of $\mathcal{P}$, and $\mathbf{B}_3$ if the tuple is in both the extension and the anti-extension of $\mathcal{P}$.

Like LP, the designated values of RM3 are both $\mathbf{T}_3$ and $\mathbf{B}_3$, so that the notions of logical consequence and theoremhood remain the same.

Theorem Proving in RM3 using Classical First-order Logic

Indirect Theorem Proving

Morgan (1976) distinguished *direct* from *indirect* theorem proving (see also (Pelletier 1991)). The underlying idea is that a logical system comes with a definition of what constitutes a proof in that system. A proof that follows such a prescription is a *direct proof* in that system. However, that is not the only way to generate a proof: an alternative is to show that a direct proof in some other system of logic guarantees the existence of a direct proof in the system of interest. This constitutes an *indirect* proof. The indirect proof is typically not a direct proof in the system of interest, e.g., the indirect proof might be ill-formed in the system of interest. It may even be the case that there is no effective way to recover a direct proof in the system of interest from the indirect proof. Nonetheless, indirect proofs are an accepted way of proving something. A common example in FOL reasoning is the conversion of a natural first-order form problem into clause normal form, and obtaining a resolution-based proof of the clause normal form (Bachmair et al. 2001). The clause normal form is not logically equivalent to the original problem, and the indirect resolution-based proof is often not a proof in the natural form, but is it accepted that the indirect proof does guarantee the existence of a natural form direct proof.

The following sections present two variants of an indirect proof method for RM3, based on translation of RM3 formulae to FOL, and use of an existing FOL reasoning systems to reason over the translated formulae. The first variant translates sentences of RM3 directly into sentences of FOL. The second variant represents the truth conditions of sentences of RM3 within FOL.

The Translational Approach

Two mutually recursive translation functions are defined for formulae in RM3: the translation *tr* and the anti-translation *atr*, resulting in formulae in FOL. Intuitively, the translation expresses the dialetheic truth of RM3 formulae (i.e., being either $\mathbf{T}_3$ or $\mathbf{B}_3$), corresponding to the extension of predicates. Similarly, the anti-translation expresses the dialetheic falsity of formulae (i.e., being either $\mathbf{B}_3$ or $\mathbf{F}_3$), corresponding to the anti-extension of predicates. The translation and anti-translation are (classically) consistent with each other: their joint truth represents the dialetheic possibility of the RM3 formula being both true and false. (The joint falsity of the translation and anti-translation does not correspond to an RM3 truth value, and is ruled out by exhaustion axioms described below.)

The translation rules are shown in Table 2. The mutual recursion terminates by translating (correspondingly, anti-translating) an RM3 atom Φ that has predicate symbol $\mathcal{P}$ and arity n, to a FOL atom Φ^+ (Φ^-) that has predicate symbol $\mathcal{P}^+$ ($\mathcal{P}^-$) also with arity n, applied to the same arguments as $\mathcal{P}$ in Φ. Φ^+ and Φ^- correspond to Φ being either $\mathbf{T}_3$ or $\mathbf{B}_3$, and $\mathbf{B}_3$ or $\mathbf{F}_3$, respectively. More precisely, if Φ^+ is $\mathbf{T}_2$ and Φ^- is $\mathbf{F}_2$ then Φ is $\mathbf{T}_3$, if Φ^+ and Φ^- are both $\mathbf{T}_2$ then Φ is $\mathbf{B}_3$, and if Φ^+ is $\mathbf{F}_2$ and Φ^- is $\mathbf{T}_2$ then Φ is $\mathbf{F}_3$,

F in RM3	$tr(F)$ in FOL
Φ	Φ^+
$\neg\phi$	$atr(\phi)$
$\phi \wedge \psi$	$tr(\phi) \wedge tr(\psi)$
$\phi \vee \psi$	$tr(\phi) \vee tr(\psi)$
$\phi \rightarrow \psi$	$\neg tr(\phi) \vee \neg atr(\psi) \vee$ $(tr(\phi) \wedge atr(\phi) \wedge tr(\psi) \wedge atr(\psi))$
$\phi \leftrightarrow \psi$	$(tr(\phi) \leftrightarrow tr(\psi)) \wedge (atr(\phi) \leftrightarrow atr(\psi))$
$\forall x A(x))$	$\forall x\, tr(A(x))$
$\exists x A(x))$	$\exists x\, tr(A(x))$
F	$atr(F)$ in FOL
Φ	Φ^-
$\neg\phi$	$tr(\phi)$
$\phi \wedge \psi$	$atr(\phi) \vee atr(\psi)$
$\phi \vee \psi$	$atr(\phi) \wedge atr(\psi)$
$\phi \rightarrow \psi$	$tr(\phi) \wedge atr(\psi)$
$\phi \leftrightarrow \psi$	$\neg(tr(\phi) \leftrightarrow tr(\psi)) \vee \neg(atr(\phi) \leftrightarrow atr(\psi)) \vee$ $(tr(\phi) \wedge atr(\phi) \wedge tr(\psi) \wedge atr(\psi))$
$\forall x A(x)$	$\exists x\, atr(A(x))$
$\exists x A(x))$	$\forall x\, atr(A(x))$

Table 2: RM3 translational approach

To ensure that every RM3 atom takes on exactly one of the three truth values ($\mathbf{T}_3$, $\mathbf{B}_3$, or $\mathbf{F}_3$), a FOL axiom of *exhaustion* is defined for each predicate $\mathcal{P}$ (with arity n) of the input:

$$\forall x_1 \cdot\cdot \forall x_n\ (\mathcal{P}^+(x_1, \cdot\cdot x_n) \vee \mathcal{P}^-(x_1, \cdot\cdot x_n))$$

The exhaustion axioms ensure that every RM3 atom takes on at least one of the three truth values, and in conjunction with the translation rules, ensure that every RM3 atom takes on exactly one of the three truth values. The exhaustion axioms also have the effect of preventing the joint falsity of the translation and anti-translation of atoms.

For a set of formulae ϕ, let $exh(\phi)$ be the set of exhaustion axioms for the predicates that occur in ϕ. Then $\phi \models_{RM3} \psi$ iff $tr(\phi) \cup exh(\phi \cup \{\psi\}) \models_{FOL} tr(\psi)$. A proof of this[1] is based on an isomorphism between RM3 interpretations of RM3 predicates $\mathcal{P}$ and classical interpretations of the corresponding FOL predicates $\mathcal{P}^+$ and $\mathcal{P}^-$, and induction over the translation rules. The isomorphism aligns the extension of $\mathcal{P}$ with the tuples that make $\mathcal{P}^+$ $\mathbf{T}_2$, and the anti-extension of $\mathcal{P}$ with the tuples that make $\mathcal{P}^-$ $\mathbf{T}_2$. As the extension and anti-extension of $\mathcal{P}$ exhaust the domain, the exhaustion axioms are satisfied by the classical model. The rules of Table 2 mimic the semantics of the RM3 connectives, so that the proof by induction is direct.

The Truth Evaluation Approach

A recursive translation function *trs* is defined for formulae in RM3, resulting in formulae in FOL. The translation function takes an RM3 formula and a target RM3 truth value (one of $\mathbf{T}_3$, $\mathbf{F}_3$, or $\mathbf{B}_3$) as arguments, and translates the formula, either directly for atoms, or recursively on the subformulae for non-atoms, to produce a FOL formula. Intuitively, the translation captures the necessary and sufficient conditions for the RM3 formula to have the target truth value.

The translation rules are shown in Table 3. The recursion terminates by translating an RM3 atom to a FOL formula that captures what it means for the atom to have the target truth value. An RM3 atom Φ that has predicate symbol $\mathcal{P}$ and arity n, is translated to a FOL formula using two FOL atoms, Φ^+ and Φ^-, which have predicate symbol $\mathcal{P}^+$ and $\mathcal{P}^-$ respectively, also with arity n, applied to the same arguments as $\mathcal{P}$ in Φ. Φ^+ and Φ^- correspond to Φ being $\mathbf{T}_2$ and $\mathbf{F}_2$ respectively.

As with the translational approach, exhaustion axioms are needed to ensure that every RM3 atom takes on at least one of the three truth values. Additionally, this approach needs *definition* axioms for each predicate $\mathcal{P}$ (with arity n) of the input, to relate the exhaustion axioms to the three truth values. The axioms introduce three new predicate symbols, $\mathcal{P}^{++}$, $\mathcal{P}^{+-}$, and $\mathcal{P}^{--}$, for each predicate symbol in the RMS problem.

$$\forall x_1 \cdot\cdot \forall x_n\ (\mathcal{P}^{++}(x_1, \cdot\cdot x_n) \leftrightarrow (\mathcal{P}^+(x_1, \cdot\cdot x_n) \wedge \neg\mathcal{P}^-(x_1, \cdot\cdot x_n)))$$
$$\forall x_1 \cdot\cdot \forall x_n\ (\mathcal{P}^{+-}(x_1, \cdot\cdot x_n) \leftrightarrow (\mathcal{P}^+(x_1, \cdot\cdot x_n) \wedge \mathcal{P}^-(x_1, \cdot\cdot x_n)))$$
$$\forall x_1 \cdot\cdot \forall x_n\ (\mathcal{P}^{--}(x_1, \cdot\cdot x_n) \leftrightarrow (\neg\mathcal{P}^+(x_1, \cdot\cdot x_n) \wedge \mathcal{P}^-(x_1, \cdot\cdot x_n)))$$

$\mathcal{P}^{++}$, $\mathcal{P}^{+-}$, and $\mathcal{P}^{--}$ correspond to $\mathbf{T}_3$, $\mathbf{B}_3$, and $\mathbf{F}_3$, respectively. The definition axioms and exhaustion axioms have the exclusive disjunction of $\mathcal{P}^{++}(x_1, \cdot\cdot x_n)$, $\mathcal{P}^{+-}(x_1, \cdot\cdot x_n)$, and $\mathcal{P}^{--}(x_1, \cdot\cdot x_n)$ as a logical consequence, so that every RM3 atom takes on exactly one of the three truth values.

For a set of formulae ϕ, let $def(\phi)$ be the set of definition axioms for the predicates that occur in ϕ. Since the designated values of RM3 are $\mathbf{T}_3$ and $\mathbf{B}_3$, define

$$des(F) \;=\; trs(F,\mathbf{T}_3) \vee trs(F,\mathbf{B}_3)$$

Then $\phi \models_{RM3} \psi$ iff $des(\phi) \cup exh(\phi \cup \{\psi\}) \cup def(\phi \cup \{\psi\}) \models_{FOL} des(\psi)$. As in the case of the translational approach, a proof of this is based on an isomorphism between

[1] Omitted here because it's too long for the paper.

F in RM3	$trs(F,\mathbf{T}_3)$ in FOL
Φ	$\Phi^+ \wedge \neg\Phi^-$
$\neg\phi$	$trs(\phi,\mathbf{F}_3)$
$\phi \wedge \psi$	$trs(\phi,\mathbf{T}_3) \wedge trs(\psi,\mathbf{T}_3)$
$\phi \vee \psi$	$trs(\phi,\mathbf{T}_3) \vee trs(\psi,\mathbf{T}_3)$
$\phi \rightarrow \psi$	$trs(\phi,\mathbf{F}_3) \vee trs(\psi,\mathbf{T}_3)$
$\phi \leftrightarrow \psi$	$(trs(\phi,\mathbf{T}_3) \wedge trs(\psi,\mathbf{T}_3) \vee$ $(trs(\phi,\mathbf{F}_3) \wedge trs(\psi,\mathbf{F}_3))$
$\forall x\ \phi$	$\forall x\ trs(\phi,\mathbf{T}_3)$
$\exists x\ \phi$	$\exists x\ trs(\phi,\mathbf{T}_3)$
F	$trs(F,\mathbf{B}_3)$
Φ	$\Phi^+ \wedge \Phi^-$
$\neg\phi$	$trs(\phi,\mathbf{B}_3)$
$\phi \wedge \psi$	$(trs(\phi,\mathbf{B}_3) \vee trs(\psi,\mathbf{B}_3)) \wedge$ $\neg trs(\phi,\mathbf{F}_3) \wedge \neg trs(\psi,\mathbf{F}_3)$
$\phi \vee \psi$	$(trs(\phi,\mathbf{B}_3) \vee trs(\psi,\mathbf{B}_3)) \wedge$ $(\neg trs(\phi,\mathbf{T}_3) \wedge \neg trs(\psi,\mathbf{T}_3))$
$\phi \rightarrow \psi$	$trs(\phi,\mathbf{B}_3) \wedge trs(\psi,\mathbf{B}_3)$
$\phi \leftrightarrow \psi$	$trs(\phi,\mathbf{B}_3) \wedge trs(\psi,\mathbf{B}_3)$
$\forall x\ \phi$	$\exists x\ trs(\phi,\mathbf{B}_3) \wedge \neg\exists x\ trs(\phi,\mathbf{F}_3)$
$\exists x\ \phi$	$\exists x\ trs(\phi,\mathbf{B}_3) \wedge \neg\exists x\ trs(\phi,\mathbf{T}_3)$
F	$trs(F,\mathbf{F}_3)$
Φ	$\neg\Phi^+ \wedge \Phi^-$
$\neg\phi$	$trs(\phi,\mathbf{T}_3)$
$\phi \wedge \psi$	$trs(\phi,\mathbf{F}_3) \vee trs(\psi,\mathbf{F}_3)$
$\phi \vee \psi$	$trs(\phi,\mathbf{F}_3) \wedge trs(\psi,\mathbf{F}_3)$
$\phi \rightarrow \psi$	$(trs(\phi,\mathbf{B}_3) \wedge trs(\psi,\mathbf{F}_3)) \vee$ $(trs(\phi,\mathbf{T}_3) \wedge (trs(\psi,\mathbf{B}_3) \vee trs(\psi,\mathbf{F}_3)))$
$\phi \leftrightarrow \psi$	$(trs(\phi,\mathbf{T}_3) \wedge (trs(\psi,\mathbf{B}_3) \vee trs(\psi,\mathbf{F}_3))) \vee$ $(trs(\phi,\mathbf{B}_3) \wedge (trs(\psi,\mathbf{T}_3) \vee trs(\psi,\mathbf{F}_3))) \vee$ $(trs(\phi,\mathbf{F}_3) \wedge (trs(\psi,\mathbf{T}_3) \vee trs(\psi,\mathbf{B}_3)))$
$\forall x\ \phi$	$\exists x\ trs(\phi,\mathbf{F}_3)$
$\exists x\ \phi$	$\forall x\ trs(\phi,\mathbf{F}_3)$

Table 3: RM3 truth evaluation approach

RM3 interpretations and classical interpretations, and induction over the translation rules.

Implementation of a Theorem Prover for RM3

In order to empirically evaluate the translation schemes, and compare the RM3 and FOL logical statuses of interesting axiom-conjecture pairs, the translation schemes have been implemented and combined with existing FOL reasoning systems to form an automated reasoning system for RM3.

The translations have been implemented in Prolog, following the translation rules of Tables 2 and 3, and adding the necessary exhaustion and uniqueness axioms. A simple shell script sends the output from the translation to one or more FOL reasoning systems. Two types of FOL reasoning systems are employed: theorem provers, to find indirect proofs of RM3 theorems, and model finders, to find countermodels of RM3 non-theorems. The shell script has command line parameters to specify which translation to use, which FOL reasoning systems to use, and to impose CPU time limits on the FOL reasoning systems. The parameters for specifying which FOL reasoning systems to use make it possible, e.g., to call a theorem prover with some time limit, and then, if no proof has been found, to call a model finder to try to find a countermodel. The system, JGRM3, is available online through the SystemOnTPTP interface (Sutcliffe 2010) at www.tptp.org/cgi-bin/SystemOnTPTP. The "A" version uses the translational approach, and the "J" variant uses the truth evaluation approach.

The theorem provers used so far are Vampire 4.1 (Kovacs and Voronkov 2013) and E 2.0 (Schulz 2013). While both are primarily designed for theorem proving, both have some countermodel finding features. Vampire can find countermodels in the form of saturations and also finite models. E can find saturations. The model finder used is Vampire 4.1, set to search only for finite models (Vampire offers various approaches to model finding). It is simple to add more FOL reasoning systems. The default configuration calls Vampire in theorem proving mode for 80% of the time limit, then swaps to Vampire in finite model finding mode for the remaining 20% of the time limit.

Comparing FOL and RM3 Problems

As was noted in the description of RM3, all theorems of RM3 are also theorems of FOL. In contrast, and of more interest, there are theorems of FOL that are not theorems of RM3. A selection of axiom-conjecture pairs is given in Table 4. All are theorems in FOL, but as indicated the in the last column of Table 4, not all are theorems of RM3. The (non-)theoremhood in RM3 theorems has been established by JGRM3, using both of the translation methods.

#	Axioms	$\models$	Conjecture	RM3?
1		$\models$	$p \vee \neg p$	Yes
2	q	$\models$	$p \rightarrow q$	No
3	$\neg p$	$\models$	$p \rightarrow q$	No
4	$\neg(p \rightarrow q)$	$\models$	p	Yes
5		$\models$	$p \rightarrow (q \vee \neg q)$	No
6		$\models$	$p \rightarrow (p \vee \neg p))$	Yes
7		$\models$	$(p \wedge \neg p) \rightarrow q$	No
8	$p \vee q, \neg p$	$\models$	q	No
9		$\models$	$(\neg p \vee q) \leftrightarrow (p \rightarrow q)$	No
10		$\models$	$(p \leftrightarrow q) \leftrightarrow ((p \wedge q) \vee (\neg p \wedge \neg q))$	No
11	$H(a)$	$\models$	$\exists x\ G(x) \rightarrow H(a)$	No
12		$\models$	$\exists x\ (G(x) \wedge \neg G(x)) \rightarrow H(b)$	No
13	$\exists x\ (G(x) \vee H(x)), \neg\exists y\ G(y)$	$\models$	$\exists z\ H(z)$	No
14	$H(a)$	$\models$	$\forall x\ (H(x) \rightarrow G(x)) \leftrightarrow$ $\forall x\ ((H(x) \wedge G(x)) \vee$ $(\neg H(x) \wedge G(a)))$	No
15		$\models$	$\neg\exists y\forall x\ (E(x,y) \leftrightarrow \neg E(x,x))$	Yes
16		$\models$	$\forall z\exists y\forall x\ (E(x,y) \leftrightarrow$ $(E(x,z) \wedge \neg E(x,x)))$ $\rightarrow \neg\exists w\forall u\ E(u,w)$	No
17		$\models$	$\neg\exists y\forall x\ (E(x,y) \leftrightarrow$ $\neg\exists z\ (E(x,z) \wedge E(z,x))$	Yes
18		$\models$	$\exists y\forall x\ (E(x,y) \leftrightarrow E(x,x)) \rightarrow$ $\neg\forall x\exists y\forall z\ (E(z,y) \leftrightarrow \neg E(z,x))$	Yes

Table 4: Example Axiom-Conjecture pairs

Many of the RM3 non-theorem cases are due directly to the differences in the truth-table for the (bi)conditional. For example, problem #2 is not an RM3 theorem because of the interpretation that sets p to $\mathbf{T}_3$ and q to $\mathbf{B}_3$, so that the axiom is designated but the conjecture is not. The contrast with classical logic is most explicit in problems #9 and #10 – these are obvious theorems of classical logic, but not theorems of RM3. RM3 has at least a tinge of "relevance" built into it, as can been seen by contrasting problems #5 and #6.

The last four problems are interesting from the point of view of the early development of set theory. Interpreting $E(x, y)$ as "x is a element of y", and thinking of all things as sets, problem #15 becomes a statement of the conclusion of Russell's Paradox: there can't be a set whose members are exactly those sets that are not elements of themselves. Russell's Paradox exists because naive set theory allows any formula to define the set of things that satisfy that formula. This classical paradox led to attempts to restrict the scope of set theory, e.g., by replacing the naive Axiom (schema) of Comprehension (which does not resolve the paradox) with the more restricted Axiom (schema) of Separation (or *Aussonderung*). This restricted axiom schema is characteristic of Zermelo set theory, and the antecedent of problem #16 is an instance of it.

Problem #16 is particularly interesting, since it isn't a theorem in RM3, and this is not *obviously* due to the different conditional. The antecedent says that for every set z there is a subset of it, y, containing just those elements x of z that are not elements of themselves. The consequent says that there is no universal set. Quine (1969, pp.37-38) pointed out that the antecedent is inconsistent with the existence of a universal set (although the two are separately consistent), so that #16 is a theorem in classical logic. However, in RM3 it is not a theorem because there is a 2-element countermodel, which JGRM3 finds. This is illustrated in Figure 1. The first domain element of the countermodel represents the universal set – $\mathcal{US}$. The second domain element represents the set of all sets that are not elements of themselves – $\mathcal{SS}$. The extension of the membership predicate E says that $\mathcal{US}$ and $\mathcal{SS}$ are both elements of themselves, and that $\mathcal{SS}$ is an element of $\mathcal{US}$. The anti-extension says $\mathcal{US}$ is a non-element of $\mathcal{SS}$, and $\mathcal{SS}$ is a non-element of itself. $\mathcal{SS}$ is the only element of $\mathcal{SS}$, and since $\mathcal{SS}$ is also a non-element of itself it is the only non-self-membered element of $\mathcal{SS}$, i.e., $\mathcal{SS}$ is the set of all sets that are not elements of themselves, and thus the antecedent of #16 is $\mathbf{T}_3$. On the other hand, everything is an element of $\mathcal{US}$, i.e., it is a universal set, and so the consequent of #16 is $\mathbf{F}_3$. Thus in RM3 #16 is not a theorem. The countermodel does not exist in classical logic because the set $\mathcal{SS}$ cannot be a element of itself and also a non-element of itself.

Table 5 shows the results of testing the translation-based system over the 1312 first-order problems without equality in v6.4.0 of the TPTP problem library (Sutcliffe 2017). The tests were done on Intel 2.4GHz CPUS with a time limit of 300s, using the default configuration of JGRM3. The translational approach generally outperforms the truth evaluation approach, but the truth evaluation approach finds slightly more FOL theorems to be RM3 non-theorems. The translational approach has many more unique solutions (problems that are not also solved by the truth evaluation approach), but the truth evaluation approach does solve some problems that the translational approach does not. Running the two approaches together results in a total of 1145 problems solved. Interesting future work would be to understand more deeply what makes a FOL theorem a non-theorem in RM3.

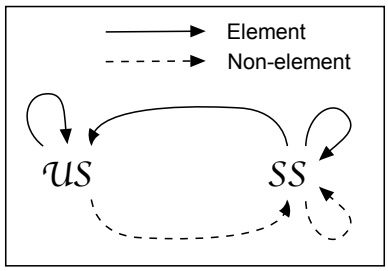

Figure 1: Countermodel for problem #16

	Probs	Trans	uniq	Truth	uniq	Total
Total	1312	1094	422	693	21	1145
FOL THM	989	508	350	158	0	508
RM3 Non-THM		363	4	378	19	382
FOL Non-THM	323	223	68	157	2	255

Table 5: Results from TPTP testing

Conclusion

Does the implementation of a reasoning system for RM3 have any real uses? Of course, advocates of RM3 will use it because the search for interesting problems that differentiate RM3 from FOL is generally quite difficult. The particular example of an interesting theorem of classical set theory might be especially important in the search for a version of dialetheic "naïve set theory". But what about more broadly?

Our framework can be adapted for other 3-valued logics by straightforward modification of the translation methods. Some of these other logics are thought to be useful for providing a formal semantics that allows for reasoning about a wide variety of natural language phenomena, such as vagueness, contingent statements about the future, presupposition failure, counterfactual reasoning, and fictional discourse. These tend to be non-paraconsistent logics where the $\mathbf{B}_3$ value means *neither true nor false*. In these truth-value "gap" logics, the exhaustion axioms are replaced by non-overlap axioms of the form:

$\forall x_1 \cdot\cdot \forall x_n \neg(\mathcal{P}^+(x_1 \cdot\cdot x_n) \wedge \mathcal{P}^-(x_1 \cdot\cdot x_n))$.

Many advocates of dialetheism wish to provide inconsistent extensions of classical first-order Peano Arithmetic. Priest (1997) advocated LP for this purpose, but the lack of a suitable conditional in LP makes it unsuited for deductive development. A variant called A3, seen perhaps as "intermediate" between LP and RM3, was used by Andrew Tedder (2015). LP and RM3 differ in two rows of the 3-valued truth table for the conditional. In one of those rows, A3's

conditional agrees with LP while in the other row it agrees with RM3.

ϕ	ψ	$(\phi \rightarrow_{LP} \psi)$	$(\phi \rightarrow_{A3} \psi)$	$(\phi \rightarrow_{RM3} \psi)$
$\mathbf{T_3}$	$\mathbf{B_3}$	$\mathbf{B_3}$	$\mathbf{B_3}$	$\mathbf{F_3}$
$\mathbf{B_3}$	$\mathbf{F_3}$	$\mathbf{B_3}$	$\mathbf{F_3}$	$\mathbf{F_3}$

Obviously this logic is amenable to both of the translation methods, and could be employed in a search for whether various arithmetic formulae of A3 are or are not theorems.

Another member of the family of logics with a small number of truth values is FDE (Belnap 1977). This logic has four truth values: three corresponding to those of LP and RM3, $\mathbf{T}_4$, $\mathbf{B}_4$, and $\mathbf{F}_4$, and additionally a fourth value $\mathbf{N}_4$ meaning *neither true nor false*, i.e., FDE drops the exhaustivity requirement. The truth values are only partially ordered: imagine a diamond with $\mathbf{T}_4$ at the top ("most true"); $\mathbf{B}_4$ and $\mathbf{N}_4$ forming the two sides of the diamond underneath $\mathbf{T}_4$; and $\mathbf{F}_4$ at the bottom ("most false"), being dominated by both $\mathbf{B}_4$ and $\mathbf{N}_4$. Sano and Omori (2014) have described a conditional for this logic, which extends that of A3. A combination of our method adapted to LP, and a variant corresponding to a 3-valued "gap" logic, will suffice. The LP part corresponds to the truth values $\mathbf{T}_4$, $\mathbf{B}_4$, and $\mathbf{F}_4$, and the gap logic's part corresponds to the truth values $\mathbf{T}_4$, $\mathbf{N}_4$, and $\mathbf{F}_4$. An argument is valid in FDE just in case the translation of its conclusion follows from the translations of its axioms, without the exhaustion axioms or the gap-logic's no-overlap axioms.

Beyond these somewhat academically useful extensions and modifications to our framework, we see real potential for leveraging paraconsistent and dialetheia-tolerating logics in the rapidly emerging application of AI techniques in the tools of modern society. It is necessary to work constructively with contradictory information, opposing but equally valid world views, argumentation forms that lead to conflicting choices, apparent inconsistencies in reality, etc. For example, RM3 might be useful for imposing constraints on large knowledge bases and ontologies, providing a non-rigid notion of consistency. Reasoning in RM3 provides one tool for AI in such scenarios.

References

Bachmair, L.; Ganzinger, H.; McAllester, D.; and Lynch, C. 2001. Resolution Theorem Proving. In Robinson, A., and Voronkov, A., eds., *Handbook of Automated Reasoning*. Elsevier Science. 19–99.

Belnap, N. 1977. A Useful Four-Valued Logic. In Dunn, J., ed., *Modern Uses of Multiple-Valued Logic*. Springer Netherlands. 5–37.

Cantini, A. 2014. Paradoxes and Contemporary Logic. In Zalta, E., ed., *The Stanford Encyclopedia of Philosophy*. Stanford University, Fall 2014 edition.

Kleene, S. 1952. *Introduction to Metamathematics*. North-Holland.

Kovacs, L., and Voronkov, A. 2013. First-Order Theorem Proving and Vampire. In Sharygina, N., and Veith, H., eds., *Proceedings of the 25th International Conference on Computer Aided Verification*, number 8044 in Lecture Notes in Artificial Intelligence, 1–35. Springer-Verlag.

Morgan, C. 1976. Methods for Automated Theorem Proving in Non-Classical Logics. *IEEE Transactions on Computers* C-25(8):852–862.

Pelletier, F. 1991. The Philosophy of Automated Theorem Proving. In Mylopolous, J., and Reiter, R., eds., *Proceedings of the 12th International Joint Conference on Artificial Intelligence*, 1039–1045. Morgan-Kaufmann.

Priest, G. 1997. Inconsistent Models of Arithmetic: Part I, Finite Models. *Journal of Philosophical Logic* 26:223–235.

Priest, G. 2006. *In Contradiction: A Study of the Transconsistent*. Oxford University Press.

Priest, G. 2008. *An Introduction to Non-Classical Logic: From If to Is*. Cambridge University Press.

Quine, W. 1969. *Set Theory and its Logic*. Harvard University Press.

Sano, K., and Omori, H. 2014. An Expansion of First-order Belnap-Dunn Logic. *Logic Journal of the IGPL* 22(3):459–481.

Schulz, S. 2013. System Description: E 1.8. In McMillan, K.; Middeldorp, A.; and Voronkov, A., eds., *Proceedings of the 19th International Conference on Logic for Programming, Artificial Intelligence, and Reasoning*, number 8312 in Lecture Notes in Computer Science, 477–483. Springer-Verlag.

Sutcliffe, G. 2010. The TPTP World - Infrastructure for Automated Reasoning. In Clarke, E., and Voronkov, A., eds., *Proceedings of the 16th International Conference on Logic for Programming, Artificial Intelligence, and Reasoning*, number 6355 in Lecture Notes in Artificial Intelligence, 1–12. Springer-Verlag.

Sutcliffe, G. 2017. The TPTP Problem Library and Associated Infrastructure. From CNF to TH0, TPTP v6.4.0. *Journal of Automated Reasoning* To appear.

Tedder, A. 2015. Axioms for Finite Collapse Models of Arithmetic. *Review of Symbolic Logic* 8:529–539.