The Thirty-First International Florida
Artificial Intelligence Research Society Conference (FLAIRS-31)

Making Belnap's "Useful Four-Valued Logic" Useful

Geoff Sutcliffe
Department of Computer Science
University of Miami

Francis Jeffry Pelletier
Department of Philosophy
University of Alberta

Allen P. Hazen
Department of Philosophy
University of Alberta

Abstract

In 1977 Nuel Belnap published two articles, "How a Computer Should Think" and "A Useful Four-Valued Logic", in which he defined a four-valued logic called FDE (First Degree Entailment). However, FDE does not allow entailments within statements, and no conditional connective is defined. As such, it is not really computationally "useful". This paper proposes conditional connectives to add to FDE, and describes the implementation of a reasoning tool for FDE with a conditional connective, with experimental results. With the addition of a conditional connective FDE starts to become truly computationally useful.

1 Background

In 1977 Nuel Belnap published two articles, "How a Computer Should Think" (Belnap 1977b) and "A Useful Four-Valued Logic" (Belnap 1977a).[1] One of the leading ideas was of a then-futuristic knowledge based system that would not only retrieve explicitly stored data, but would also reason and deduce consequences of the stored data. A further idea was that such a knowledge base might be given contradictory data to store, and that there might be topics for which no data is stored. For example, a knowledge base about characters in a TV series might be used to determine what the script could include (e.g., live characters require words and placement, dead characters require only placement). However, characters could be both alive and not alive (e.g, the undead), or not in the knowledge base at all (e.g., not yet in the series). Reasoning about these situations in classical logic is problematic - contradictions allow any consequence to be deduced (e.g., the existence of the undead leads to the conclusion that dead characters need words), and in the absence of knowledge tautologies can be deduced (e.g., non-existent characters are alive or not alive, and hence require placement). Belnap thought that this situation called for a four-valued logic, where the truth-values are **T**rue, **F**alse, **B**oth, and **N**either. This logic was given the name FDE (First Degree Entailment).

Belnap envisaged the four truth values of FDE in two lattice forms, the "Truth Diamond" and the "Information Diamond", shown in Figure 1. The Truth Diamond represents the amount of *truth* in the four truth values, with **T** having the most (only true) and **F** the least (only false). **B** and **N** are between the two extremes of **T** and **F**, with different ways of balancing their true and false parts, and therefore incomparable amounts of truth. In contrast, the Information Diamond represents the amounts of *information* in the four truth values, with **B** having the most (both true and false) and **N** the least (neither true nor false). **T** and **F** are between the two extremes of **B** and **N** with different types, and therefore incomparable amounts, of information.

Figure 1: The Truth and Information Diamonds

Three logical connectives were provided for FDE: conjunction ($\wedge$), disjunction ($\vee$), and negation ($\neg$). The truth value of a conjunction (disjunction) is the meet (join) of its conjuncts (disjuncts) in the lattice, and negation inverts the lattice order. Table 1 shows these as truth matrices. As usual, the truth values are divided into those that are *designated* – the values that "true" statements should have (like being **T** in classical logic), and those that are *undesignated*. Logical truths are those that are always designated regardless of the truth values of their atomic components, and are the formulae that a reasoning tool should be able to prove. In FDE **T** and **B** are designated, so that **N** and **F** are undesignated.

$\wedge$	T	B	N	F
T	T	B	N	F
B	B	B	F	F
N	N	F	N	F
F	F	F	F	F

$\vee$	T	B	N	F
T	T	T	T	T
B	T	B	T	B
N	T	T	N	N
F	T	B	N	F

$\neg$	
T	F
B	B
N	N
F	T

Table 1: Truth matrices for FDE's connectives

FDE was formulated without a conditional connective, and instead meta-theoretically allowed entailments of the

[1]Most easily accessed as the merged version, (Belnap 1992).

form $\Gamma \vdash \varphi$ among statements. That way entailments could never become embedded - it would be ill-formed to say anything like $(\Gamma \vdash \varphi) \vdash (\Delta \vdash \psi)$. Thus there is a notion of just *first-degree* entailments, rather than a connective that accepts arbitrary sentences as its arguments (Dunn 1976). This makes it impossible to state that some formula follows from given premises, and the ability to reason and deduce consequences is restricted. (Rodrigues and Russo 1998) provide a partial solution by employing a first-order representation of FDE in which the meta-linguistic entailment $\vdash$ is represented by a special conditional connective that can occur at most once in any statement.

A second weakness of FDE is that it is not *functionally complete* – only a subset of the $4^{(4^2)}$ (= 4,294,967,296) possible binary relations among the four values is expressible in the language of $\wedge$, $\vee$, and $\neg$. For example, there is no truth function that starts with all the atomic letters being assigned **N** and results in an expression that has a value other than **N**.

The lack of a suitable conditional connective and the lack of functional completeness need to be addressed in order for FDE to assume the "computationally useful" role envisaged by Belnap. This paper proposes two conditional connectives to add to FDE, which makes it possible to exploit the full range of the inferences that should be allowed in a four-valued logic like FDE, so that it starts to become truly computationally useful. The issue of functional completeness will be addressed in future work.

2 What is a "Useful" Conditional?

In classical logic it is normal to define a conditional connective in terms of $\vee$ and $\neg$

$$(\varphi \rightarrow \psi) =_{df} (\neg\varphi \vee \psi)$$

Using that definition in FDE yields a conditional connective with the truth-matrix shown in Table 2 (the antecedent value is in the left column, the consequent value is in the top row).

$\rightarrow$	T	B	N	F
T	T	B	N	F
B	T	B	T	B
N	T	T	N	N
F	T	T	T	T

Table 2: Truth matrix for classical definition of $\rightarrow$ in FDE

However, modus ponens (MP) – the central principle of reasoning and chaining results together – fails in FDE with this conditional connective. Let $[\![\varphi]\!] = TV$ express that the truth-value of the formula φ is TV. Then, for example, if $[\![\varphi]\!] = \mathbf{B}$ and $[\![\psi]\!] = \mathbf{N}$ modus ponens fails, because $[\![\varphi]\!]$ and $[\![\varphi \rightarrow \psi]\!]$ are designated but $[\![\psi]\!]$ is not. (Correspondingly, unit resolution also fails in this logic: If $[\![\varphi]\!] = \mathbf{B}$ and $[\![\psi]\!] = \mathbf{F}$, then $[\![\neg\varphi]\!] = \mathbf{B}$ and $[\![\varphi \vee \psi]\!] = \mathbf{B}$. Both are designated, but the unit resolvent $[\![\psi]\!] = \mathbf{F}$ is not.) Thus a different way of defining a conditional connective must be used. The overarching consideration is that the conditional connective must support MP, i.e., whenever the antecedent and the conditional have designated values, the consequent must have a designated value.

One way to ensure that MP holds is to emphasize the classical aspect of FDE. In the cases when the antecedent is designated, MP holds naturally if the value of the consequent is assigned to the conditional. This is the case in classical logic ($\mathbf{T} \rightarrow \mathbf{T} = \mathbf{T}$ and $\mathbf{T} \rightarrow \mathbf{F} = \mathbf{F}$), and the principle can be adopted here (e.g., $\mathbf{B} \rightarrow \mathbf{N} = \mathbf{N}$). In the cases when the antecedent is undesignated MP holds vacuously, and the value of the conditional does not matter, regardless of the value of the consequent. In the classical two-valued case, **T** is the value that is assigned to the conditional when the antecedent is **F**. The corresponding idea in FDE is to assign **T** or **B** to the conditional when the antecedent is **N** or **F**. The conditional connective described in (Avron 1991), and considered in (Hazen and Pelletier 2017), assigns **T** to the conditional whenever the antecedent is undesignated. These classicality considerations define a suitable conditional connective for FDE, called $\rightarrow_{cmi}$ (for **c**lassical **m**aterial **i**mplication). Its truth matrix is given in Table 3.

$\rightarrow_{cmi}$	T	B	N	F
T	T	B	N	F
B	T	B	N	F
N	T	T	T	T
F	T	T	T	T

Table 3: Truth matrix for $\rightarrow_{cmi}$

It is noteworthy that FDE with the connectives $\wedge, \vee, \neg$, but restricted to the three truth values **T**, **N**, and **F**, is the "strong Kleene 3-valued logic" K3 (Kleene 1952, §64). Similarly, FDE restricted to **T**, **B**, and **F**, is the "logic of paradox" LP (Priest 1979). The investigation in (Hazen and Pelletier 2017) shows that adding $\rightarrow_{cmi}$ to K3 generates a logic "synonymous" (Pelletier and Urquhart 2003) with Łukasiewicz's three-valued logic Ł3 (in which the third value is **N**) (Łukasiewicz 1920; Łukasiewicz and Tarski 1930). Similarly, adding $\rightarrow_{cmi}$ to LP generates the logic A3 (in which the third value is **B**) (Avron 1991; Tedder 2015). It is due to the ease of proving these sorts of meta-theoretic results that the $\rightarrow_{cmi}$ conditional was chosen in (Hazen and Pelletier 2017) to add to FDE, and why it is considered here. A related logic is RM3 (Pelletier, Sutcliffe, and Hazen 2017), which like A3 can be built by adding a conditional connective to LP. RM3's conditional connective is different to $\rightarrow_{cmi}$, defined by the contrapositive of $\rightarrow_{cmi}$

$$(\phi \rightarrow_{rm3} \psi) =_{df} ((\phi \rightarrow_{cmi} \psi) \wedge (\neg\psi \rightarrow_{cmi} \neg\phi)).$$

It turns out that FDE$^{\rightarrow}$(FDE with either of the two conditionals proposed in this paper) is a sublogic of A3, which is a sublogic of FOL, i.e., the FDE$^{\rightarrow}$ theorems are a subset of theorems of A3, which are a subset of the theorems of FOL. The different conditional connective of RM3 places it apart from A3 and FDE$^{\rightarrow}$, but RM3 is also a sublogic of FOL. A comparison between reasoning in RM3, A3, and FDE$^{\rightarrow}$is given in Section 4.

The decisions made for the $\rightarrow_{cmi}$ connective were based on notions adopted from classical logic. These formed some *constraints* on possible values for the connective values, from which intuitively desirable values were selected. This principle of using constraints was further developed in this research, to the point where there were only four choices left, for which "obvious" decisions were made. This resulted in a second (but similar) conditional connective for FDE, called $\rightarrow_{con}$ (for **con**strained). The following are the constraints that were used to produce $\rightarrow_{con}$, in our perceived order of plausibility.

1. Modus Ponens: If the antecedent is designated, and the consequent is undesignated, then the conditional is undesignated.
2. Classicality: If the antecedent and consequent are either **T** or **F** then the conditional must agree with classical logic. Thus
$$\mathbf{T} \rightarrow_{con} \mathbf{T} = \mathbf{T} \qquad \mathbf{T} \rightarrow_{con} \mathbf{F} = \mathbf{F}$$
$$\mathbf{F} \rightarrow_{con} \mathbf{T} = \mathbf{T} \qquad \mathbf{F} \rightarrow_{con} \mathbf{F} = \mathbf{T}$$
3. False Implies True: If the antecedent is **F** or the consequent is **T** then the conditional is **T**. Thus
$$\mathbf{F} \rightarrow_{con} \mathbf{B} = \mathbf{T} \qquad \mathbf{F} \rightarrow_{con} \mathbf{N} = \mathbf{T}$$
$$\mathbf{B} \rightarrow_{con} \mathbf{T} = \mathbf{T} \qquad \mathbf{N} \rightarrow_{con} \mathbf{T} = \mathbf{T}$$
4. Undesignated Antecedent: If the antecedent is undesignated then the conditional is designated. Thus
$$\mathbf{N} \rightarrow_{con} \mathbf{B} = \mathbf{T} \textit{ or } \mathbf{B} \qquad \mathbf{N} \rightarrow_{con} \mathbf{N} = \mathbf{T} \textit{ or } \mathbf{B}$$
$$\mathbf{N} \rightarrow_{con} \mathbf{F} = \mathbf{T} \textit{ or } \mathbf{B}$$
5. Designated Antecedent in a Diamond: If the antecedent is designated, and above or equal to the consequent in the Truth Diamond or the Information Diamond, then the value of the conditional is that of the consequent. Thus
$$\mathbf{T} \rightarrow_{con} \mathbf{B} = \mathbf{B} \qquad \mathbf{T} \rightarrow_{con} \mathbf{N} = \mathbf{N}$$
$$\mathbf{B} \rightarrow_{con} \mathbf{B} = \mathbf{B} \qquad \mathbf{B} \rightarrow_{con} \mathbf{N} = \mathbf{N}$$
$$\mathbf{B} \rightarrow_{con} \mathbf{N} = \mathbf{N}$$
6. Non-equivalence: With the biconditional $\phi \leftrightarrow \psi$ defined as $(\phi \rightarrow \psi) \wedge (\psi \rightarrow \phi)$, for distinct truth values TV_1 and TV_2, $TV_1 \leftrightarrow TV_2 \neq \mathbf{T}$. Thus
$$\mathbf{N} \rightarrow_{con} \mathbf{F} = \mathbf{B}$$

These constraints define all values for $\rightarrow_{con}$ except $\mathbf{N} \rightarrow_{con} \mathbf{B}$ and $\mathbf{N} \rightarrow_{con} \mathbf{N}$, each of which can be **T** *or* **B**, i.e., there are four possibilities.[2] For $\mathbf{N} \rightarrow_{con} \mathbf{N}$ the obvious choice is **T**, as that makes $\mathbf{N} \leftrightarrow_{con} \mathbf{N} = \mathbf{T}$. Then believing that $\mathbf{N} \rightarrow_{con} \mathbf{B}$ should be different from $\mathbf{N} \rightarrow_{con} \mathbf{N}$, $\mathbf{N} \rightarrow_{con} \mathbf{B} = \mathbf{B}$. The resultant matrix for $\rightarrow_{con}$ is given in Table 4. There are only two differences from $\rightarrow_{cmi}$, for which $\mathbf{N} \rightarrow \mathbf{B}$ and $\mathbf{N} \rightarrow \mathbf{F}$ are **T**. These are boldfaced in Table 4.

$\rightarrow_{con}$	T	B	N	F
T	T	B	N	F
B	T	B	N	F
N	T	**B**	T	**B**
F	T	T	T	T

Table 4: Truth matrix for $\rightarrow_{con}$

Constraints 2-6 alone ensure that MP is satisfied, so that the Modus Ponens constraint is not necessary for ensuring that the conditional connective is MP-compliant. Some of the constraints confirm what earlier constraints had already required, e.g., Classicality requires $\mathbf{F} \rightarrow_{con} \mathbf{T} = \mathbf{T}$, which is also required by the subsequent constraints False Implies True and Undesignated Antecedent.

There are also other properties that a good conditional connective should have, and do hold for $\rightarrow_{con}$. These include:

- Deduction Theorem: If the fact that the antecedent is designated can be used to prove that the consequent is designated, then it is possible to prove that the conditional is designated.
- Chained Implication: If $\phi \rightarrow \psi$ and $\psi \rightarrow \sigma$ are designated then $\phi \rightarrow \sigma$ is designated.
- Contraction: $(\phi \rightarrow (\phi \rightarrow \psi)) \rightarrow (\phi \rightarrow \psi)$ is designated.
- Thinning: $\phi \rightarrow (\psi \rightarrow \phi)$ is designated.
- Designated Antecedent: If the antecedent is designated then the value of the conditional is that of the consequent (see, e.g., (Arieli and Avron 1998)).

Some properties of a good conditional connective do not hold for $\rightarrow_{con}$, e.g.,

- Contraposition: $(\phi \rightarrow \psi) \leftrightarrow (\neg\psi \rightarrow \neg\phi)$

The process of producing the final four possible conditional connectives was automated by encoding $\wedge$, $\vee$, $\neg$, the diamonds, and the constraints in first-order logic. The encoding was passed to the automated reasoning system Vampire 4.2 (Kovacs and Voronkov 2013), configured to find a finite model. The finite model produced includes the values for a conditional connective. Iteratively, the negation of each such conditional connective found was added to the encoding so that the next run would produce a different conditional connective. The encoding plus the negations of the four possible conditional connectives was found to be unsatisfiable, indicating there are no more possible conditional connectives that meet the constraints.

Adding $\rightarrow_{cmi}$ ($\rightarrow_{con}$) to FDE produces the logic $\text{FDE}^{\rightarrow_{cmi}}$ ($\text{FDE}^{\rightarrow_{con}}$). Even though $\rightarrow_{cmi}$ and $\rightarrow_{con}$ and their corresponding biconditionals are different, it is heartening to note that they are very similar: The conditionals differ only when the antecedent is **N**, and the consequent is **B** or **F**. The biconditionals differ only when one side is **N** and the other side is **B** or **F**. In all cases the difference is only in which of the designated or undesignated values the conditional or biconditional produces

$\mathbf{N} \rightarrow_{cmi} \mathbf{B} = \mathbf{T}$	$\mathbf{N} \rightarrow_{con} \mathbf{B} = \mathbf{B}$	(designated)
$\mathbf{N} \rightarrow_{cmi} \mathbf{F} = \mathbf{T}$	$\mathbf{N} \rightarrow_{con} \mathbf{F} = \mathbf{B}$	(designated)
$\mathbf{N} \leftrightarrow_{cmi} \mathbf{B} = \mathbf{N}$	$\mathbf{N} \leftrightarrow_{con} \mathbf{B} = \mathbf{F}$	(undesignated)
$\mathbf{N} \leftrightarrow_{cmi} \mathbf{F} = \mathbf{T}$	$\mathbf{N} \leftrightarrow_{con} \mathbf{F} = \mathbf{B}$	(designated)

This demonstrates that the meta-theoretic bases for $\rightarrow_{cmi}$ closely mirror the constraints on a good conditional that produced $\rightarrow_{con}$. In Sections 3 and 4 the two logics $\text{FDE}^{\rightarrow_{cmi}}$ and $\text{FDE}^{\rightarrow_{con}}$ are referred to generically as $\text{FDE}^{\rightarrow}$ when the results and comments apply to both variants.

Although $\rightarrow_{cmi}$ and $\rightarrow_{con}$ are the same in terms of designation, $\text{FDE}^{\rightarrow_{cmi}}$ and $\text{FDE}^{\rightarrow_{con}}$ have different theorems. This can be seen in Problems 4 and 11 of Table 6 in Section 4, which might make $\text{FDE}^{\rightarrow_{con}}$ unacceptable to some classical

[2] The second author does not subscribe to Non-equivalence, so for him there are eight possibilities.

logicians. Another stark difference is between their biconditionals for $\mathbf{N} \leftrightarrow \mathbf{F}$. While FDE$^{\to_{cmi}}$ claims that **N** and **F** are truly equivalent, FDE$^{\to_{con}}$ says **N** and **F** are both equivalent and non-equivalent. This leads to counterintuitive possibilities in FDE$^{\to_{cmi}}$. For example, a paramodulation-like inference rule that replaces a formula by a truly equivalent formula could replace a formula that is **N** by one that is **F**. This counterintuitive replacement would not occur in FDE$^{\to_{con}}$.

3 Implementation

In (Pelletier, Sutcliffe, and Hazen 2017) two translations from the logic RM3 to classical first-order logic (FOL) were presented, providing an indirect theorem proving method for RM3. The "truth evaluation" translation has been adopted to produce an indirect theorem proving method for FDE$^{\to}$.

A recursive translation function *trs* is defined for formulae in FDE$^{\to}$, resulting in formulae in FOL. The translation function takes an FDE$^{\to}$ formula and a target FDE$^{\to}$ truth value (one of **T**, **B**, **N**, or **F**) 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 FDE$^{\to}$ formula to have the target truth value. The translation rules are shown in Table 5. The recursion terminates by translating an FDE$^{\to}$ atom to a FOL atom. Equality is treated classically, so that for a target truth value of **B** or **N** an equality atom is translated to the FOL constant **F**. Non-equality atoms are translated to a FOL atom that captures what it means for the atom to have the target truth value. A FDE$^{\to}$ atom Φ that has predicate symbol $\mathcal{P}$ and arity n, is translated to a FOL atom with the predicate symbol $\mathcal{P}^T$ or $\mathcal{P}^B$ or $\mathcal{P}^N$ or $\mathcal{P}^F$, corresponding to the target truth value, also with arity n, applied to the same arguments as $\mathcal{P}$ in Φ. *Definition Axioms* are added to relate each predicate symbol $\mathcal{P}^T$, $\mathcal{P}^B$, $\mathcal{P}^N$, and $\mathcal{P}^F$ to atoms that correspond to the two truth values **T** and **F** of FOL. The axioms introduce two new predicate symbols, $\mathcal{P}^{cT}$ and $\mathcal{P}^{cF}$ (for **c**lassical **T**rue and **F**alse) for each predicate symbol $\mathcal{P}$ in the FDE$^{\to}$ problem. The axioms are:

$$\forall x_1 \cdot\cdot \forall x_n\ (\mathcal{P}^T(x_1, \cdot\cdot x_n) \leftrightarrow (\mathcal{P}^{cT}(x_1, \cdot\cdot x_n) \land \neg \mathcal{P}^{cF}(x_1, \cdot\cdot x_n)))$$
$$\forall x_1 \cdot\cdot \forall x_n\ (\mathcal{P}^B(x_1, \cdot\cdot x_n) \leftrightarrow (\mathcal{P}^{cT}(x_1, \cdot\cdot x_n) \land \mathcal{P}^{cF}(x_1, \cdot\cdot x_n)))$$
$$\forall x_1 \cdot\cdot \forall x_n\ (\mathcal{P}^N(x_1, \cdot\cdot x_n) \leftrightarrow (\neg \mathcal{P}^{cT}(x_1, \cdot\cdot x_n) \land \neg \mathcal{P}^{cF}(x_1, \cdot\cdot x_n)))$$
$$\forall x_1 \cdot\cdot \forall x_n\ (\mathcal{P}^F(x_1, \cdot\cdot x_n) \leftrightarrow (\neg \mathcal{P}^{cT}(x_1, \cdot\cdot x_n) \land \mathcal{P}^{cF}(x_1, \cdot\cdot x_n)))$$

Finally, *Exhaustion Axioms* are added to enforce that each FDE$^{\to}$ atom takes on exactly one of the four truth values. The axioms are:

$$\forall x_1 \cdot\cdot \forall x_n\ (\mathcal{P}^T(x_1, \cdot\cdot x_n) \lor \mathcal{P}^B(x_1, \cdot\cdot x_n)\ \lor$$
$$\mathcal{P}^N(x_1, \cdot\cdot x_n) \lor \mathcal{P}^F(x_1, \cdot\cdot x_n))$$

The exclusive disjunction of the four disjuncts follows from these axioms and the definition axioms.

For a set of formulae ϕ, let $def(\phi)$ be the set of definition axioms, and $exh(\phi)$ the set of exhaustion axioms, for the predicate symbols that occur in ϕ. Since the designated values of FDE$^{\to}$ are **T** and **B**, define

$$des(\phi)\ =\ trs(\phi,\mathbf{T}) \lor trs(\phi,\mathbf{B})$$

For a problem $\phi \vDash \psi$ define

$$trans(\phi)\ =\ des(\phi) \cup exh(\phi \cup \{\psi\}) \cup def(\phi \cup \{\psi\})$$

Then $\phi\ \ \vDash_{FDE^{\to}}\ \ \psi$ iff $trans(\phi)\ \ \vDash_{FOL}\ \ des(\psi)$. A

F	$trs(F,\mathbf{T})$
$\phi = \psi$	$\phi = \psi$
Φ	Φ^T
$\neg\phi$	$trs(\phi,\mathbf{F})$
$\phi \land \psi$	$trs(\phi,\mathbf{T}) \land trs(\psi,\mathbf{T})$
$\phi \lor \psi$	$trs(\phi,\mathbf{T}) \lor trs(\psi,\mathbf{T}) \lor$ $(trs(\phi,\mathbf{B}) \land trs(\psi,\mathbf{N}) \lor (trs(\psi,\mathbf{B}) \land trs(\phi,\mathbf{N})$
$\phi \to_{cmi} \psi$	$trs(\phi,\mathbf{F}) \lor trs(\psi,\mathbf{T}) \lor trs(\phi,\mathbf{N})$
$\phi \to_{con} \psi$	$trs(\phi,\mathbf{F}) \lor trs(\psi,\mathbf{T}) \lor (trs(\phi,\mathbf{N}) \land trs(\psi,\mathbf{N}))$
$\phi \leftrightarrow_{cmi} \psi$	$(trs(\phi,\mathbf{T}) \land trs(\psi,\mathbf{T}) \lor$ $((trs(\phi,\mathbf{N}) \lor trs(\phi,\mathbf{F})) \land (trs(\psi,\mathbf{N}) \lor trs(\psi,\mathbf{F})))$
$\phi \leftrightarrow_{con} \psi$	$(trs(\phi,\mathbf{T}) \land trs(\psi,\mathbf{T}))\ \lor (trs(\phi,\mathbf{N}) \land trs(\psi,\mathbf{N})) \lor$ $(trs(\phi,\mathbf{F}) \land trs(\psi,\mathbf{F}))$
$\forall x\ \phi$	$\forall x\ trs(\phi,\mathbf{T})$
$\exists x\ \phi$	$\exists x\ trs(\phi,\mathbf{T})$
F	$trs(F,\mathbf{B})$
$\phi = \psi$	$\mathbf{F}$
Φ	Φ^B
$\neg\phi$	$trs(\phi,\mathbf{B})$
$\phi \land \psi$	$(trs(\phi,\mathbf{B}) \land trs(\psi,\mathbf{B})) \lor (trs(\phi,\mathbf{B}) \land trs(\psi,\mathbf{T}) \lor$ $(trs(\phi,\mathbf{F}) \land \neg trs(\psi,\mathbf{B}))$
$\phi \lor \psi$	$(trs(\phi,\mathbf{B}) \land trs(\psi,\mathbf{B})) \lor (trs(\phi,\mathbf{B}) \land trs(\psi,\mathbf{F})) \lor$ $(trs(\phi,\mathbf{F}) \land trs(\psi,\mathbf{B}))$
$\phi \to_{cmi} \psi$	$(trs(\phi,\mathbf{T}) \lor trs(\phi,\mathbf{B}) \land trs(\psi,\mathbf{B})$
$\phi \to_{con} \psi$	$((trs(\phi,\mathbf{T}) \lor trs(\phi,\mathbf{B}) \lor trs(\phi,\mathbf{N})) \land trs(\psi,\mathbf{B})) \lor$ $(trs(\phi,\mathbf{N}) \land trs(\psi,\mathbf{F}))$
$\phi \leftrightarrow_{cmi} \psi$	$(trs(\phi,\mathbf{T}) \land trs(\psi,\mathbf{B})) \lor (trs(\phi,\mathbf{B}) \land trs(\psi,\mathbf{T})) \lor$ $(trs(\phi,\mathbf{B}) \land trs(\psi,\mathbf{B}))$
$\phi \leftrightarrow_{con} \psi$	$(trs(\phi,\mathbf{T}) \land trs(\psi,\mathbf{B})) \lor (trs(\phi,\mathbf{B}) \land trs(\psi,\mathbf{T})) \lor$ $(trs(\phi,\mathbf{B}) \land trs(\psi,\mathbf{B})) \lor (trs(\phi,\mathbf{N}) \land trs(\psi,\mathbf{F})) \lor$ $(trs(\phi,\mathbf{F}) \land trs(\psi,\mathbf{N}))$
$\forall x\ \phi$	$\exists x\ trs(\phi,\mathbf{B}) \land \neg\exists x\ trs(\phi,\mathbf{F})$
$\exists x\ \phi$	$\exists x\ trs(\phi,\mathbf{B}) \land \neg\exists x\ trs(\phi,\mathbf{T})$
F	$trs(F,\mathbf{N})$
$\phi = \psi$	$\mathbf{F}$
Φ	Φ^N
$\neg\phi$	$trs(\phi,\mathbf{N})$
$\phi \land \psi$	$(trs(\phi,\mathbf{N}) \land trs(\psi,\mathbf{N})) \lor (trs(\phi,\mathbf{T}) \land trs(\psi,\mathbf{N})) \lor$ $(trs(\phi,\mathbf{N}) \land trs(\psi,\mathbf{T}))$
$\phi \lor \psi$	$(trs(\phi,\mathbf{N}) \land trs(\psi,\mathbf{N})) \lor (trs(\phi,\mathbf{N}) \land trs(\psi,\mathbf{F})) \lor$ $(trs(\phi,\mathbf{F}) \land trs(\psi,\mathbf{N}))$
$\phi \to \psi$	$(trs(\phi,\mathbf{T}) \lor trs(\phi,\mathbf{B})) \land trs(\psi,\mathbf{N})$
$\phi \leftrightarrow_{cmi} \psi$	$(trs(\phi,\mathbf{N}) \land (trs(\psi,\mathbf{T}) \lor trs(\psi,\mathbf{B}))) \lor$ $(trs(\psi,\mathbf{N}) \land (trs(\phi,\mathbf{T}) \lor trs(\phi,\mathbf{B})))$
$\phi \leftrightarrow_{con} \psi$	$(trs(\phi,\mathbf{T}) \land trs(\psi,\mathbf{N})) \lor (trs(\phi,\mathbf{N}) \land trs(\psi,\mathbf{T}))$
$\forall x\ \phi$	$\exists x\ trs(\phi,\mathbf{N}) \land \neg\exists x\ trs(\phi,\mathbf{F})$
$\exists x\ \phi$	$\exists x\ trs(\phi,\mathbf{N}) \land \neg\exists x\ trs(\phi,\mathbf{T})$
F	$trs(F,\mathbf{F})$
$\phi = \psi$	$\phi \neq \psi$
Φ	Φ^F
$\neg\phi$	$trs(\phi,\mathbf{T})$
$\phi \land \psi$	$trs(\phi,\mathbf{F}) \lor trs(\psi,\mathbf{F}) \lor (trs(\phi,\mathbf{B}) \land trs(\psi,\mathbf{N})) \lor$ $(trs(\phi,\mathbf{N}) \land trs(\psi,\mathbf{B}))$
$\phi \lor \psi$	$trs(\phi,\mathbf{F}) \land trs(\psi,\mathbf{F})$
$\phi \to \psi$	$(trs(\phi,\mathbf{T}) \lor trs(\phi,\mathbf{B})) \land trs(\psi,\mathbf{F})$
$\phi \leftrightarrow_{cmi} \psi$	$((trs(\phi,\mathbf{T}) \lor trs(\phi,\mathbf{B})) \land trs(\psi,\mathbf{F})) \lor$ $((trs(\psi,\mathbf{T}) \lor trs(\psi,\mathbf{B})) \land trs(\phi,\mathbf{F}))$
$\phi \leftrightarrow_{con} \psi$	$((trs(\phi,\mathbf{T}) \lor trs(\phi,\mathbf{B})) \land trs(\psi,\mathbf{F})) \lor$ $((trs(\psi,\mathbf{T}) \lor trs(\psi,\mathbf{B})) \land trs(\phi,\mathbf{F})) \lor$ $(trs(\phi,\mathbf{B}) \land trs(\psi,\mathbf{N})) \lor (trs(\phi,\mathbf{N}) \land trs(\psi,\mathbf{B}))$
$\forall x\ \phi$	$\exists x\ trs(\phi,\mathbf{F})$
$\exists x\ \phi$	$\forall x\ trs(\phi,\mathbf{F})$

Table 5: FDE$^{\to}$ truth evaluation translation

theorem prover for FDE is simply implemented by submitting $trans(\phi) \vDash_{FOL} des(\psi)$ to a FOL theorem prover such as Vampire. The implementation can be used online in the SystemOnTPTP interface, at `www.tptp.org/cgi-bin/SystemOnTPTP`. The system is called `JGXYZ FDEJ-01` for FDE$^{\to_{cmi}}$, and `JGXYZ FDEG-01` for FDE$^{\to_{con}}$

4 Results

The implementation of FDE$^{\to}$ (both the FDE$^{\to_{cmi}}$ and FDE$^{\to_{con}}$ variants) has been tested on a set of problems, mostly from (Pelletier, Sutcliffe, and Hazen 2017) where they were used to study the differences between reasoning in FOL and RM3. They have been reused here, with a couple of new problems, to gain insights into the differences between reasoning in FOL, RM3, A3, and FDE$^{\to}$. Table 6 gives the results of the testing. All the problems are valid in FOL. The (translations of the) problems marked "Yes" were proved with Vampire, and the entries marked "No" had countermodels generated by Vampire in finite model finding mode.

There are differences in the provability-status between RM3 and A3, as expected. Because of their different conditional connectives, some non-theorems of RM3 are theorems of A3, as seen in Problems 2, 5, and 12. In Problem 2, if $[\![q]\!]$ is **B** and $[\![p]\!]$ is **T** then $[\![p \to q]\!]$ is **F** in RM3 (hence not a theorem) but **B** in A3 (hence a theorem). In Problem 5, $[\![q \vee \neg q]\!]$ = **T** *or* **B**, i.e., valid in both RM3 and A3 (as shown by Problem 1). However, if $[\![p]\!]$ is **T** and $[\![q \vee \neg q]\!]$ is **B**, then the conjecture is **F** in RM3 but **B** in A3.

In FDE$^{\to}$ the additional truth value **N** results in some RM3 and A3 theorems becoming non-theorems, as seen in Problems 1 and 5. In Problem 1, if $[\![p]\!]$ is **N** then $[\![p \vee \neg p]\!]$ = **N**, and hence undesignated (not a theorem) in FDE$^{\to}$. In Problem 5, if $[\![p]\!]$ is **T** and $[\![q]\!]$ is **N** then $[\![p \to (q \vee \neg q)]\!]$ = **N**. These assignments can't happen in RM3 or A3, since there is no **N** value. Problems 4 and 11 show that neither FDE$^{\to_{cmi}}$ nor FDE$^{\to_{con}}$ is a sublogic of the other. In Problem 4, if $[\![p]\!]$ is **N** and $[\![q]\!]$ is **B** then in FDE$^{\to_{cmi}}$ it's a theorem, but in FDE$^{\to_{con}}$ $[\![\neg(p \to q)]\!]$ = **B** - the assignment is a countermodel and it's not a theorem. In Problem 11, if $[\![p]\!]$ is **N** and $[\![q]\!]$ is **F** then in FDE$^{\to_{con}}$ it's a theorem, but in FDE$^{\to_{cmi}}$ the formula is **N** and the assignment is a countermodel.

Problems 16-19 are interesting both from a historical and also a contemporary point of view of the foundations of mathematics. They represent some of the motivating claims that drove the modern development of axiomatic set theory and mathematics. Read the relation $E(x, y)$ as saying that x *is an element of (set)* y. Then each formula represents a crucial part of the various paradoxes of set theory. For example, Russell's paradox is in part captured by Problem 16, which says that there cannot be a set (y) all of whose members (x) are not members of themselves. Problem 17 represents a further paradox that is involved with one of the attempts to provide a solution to the Russell paradox, saying that if set membership is restricted so that a set being defined must be a subset of an already-established set, then there cannot be a universal set (a set that contains all sets). Problem 18 says there can't be a set that contains "circularly-contained" sets. Problem 19 says that if there is set that contains all and only those sets that are members of themselves, then not every set can have a complement. From a modern point of view, there is a small (but vocal!) movement to re-establish the validity of a "naïve comprehension" principle, the presumption of which is generally thought to be the source of these and other paradoxes. If a dialetheic logic such as RM3 or A3 or FDE$^{\to}$ were to be adopted, then perhaps the naïve comprehension principle could be retained. In (Pelletier, Sutcliffe, and Hazen 2017) we were excited to discover that RM3 did not prove Problem 17, and that a counter-model was found in terms of the three-valued logic. However, Table 6 shows that Problems 16, 18, and 19 remain provable in RM3 and A3, showing that adding the third value **B** does not solve all the difficulties of naïve set theory. However, those problems are invalid in FDE$^{\to}$, which might indicate that FDE$^{\to}$ could be a suitable basis for naïve set theory. Some of the steps were taken by (Tedder 2015) in the context of A3, and perhaps those same constructions could be adapted to FDE$^{\to}$.

Problem 20 was not one of the problems considered for the RM3 study because it involves identity, which was not a part of RM3 as developed in (Pelletier, Sutcliffe, and Hazen 2017). Classical identity has since been added to that system, and the result is given in Table 6. This problem considers the possibility of an infinite descending chain of memberships within a given class (set) a, that is, an infinite sequence $\cdots \in b_n \in \cdots \in b_2 \in b_1 \in a$. A set is grounded if it does not exhibit such an infinite chain, and intuitively, all sets are grounded. (Kalish and Montague 1964, p.226) present this as a homework problem within first order logic. The argument concerning the notion of a grounded set goes like this: If the set of all grounded sets, call it g, is grounded, then $\cdots \in g \in \cdots \in g$ and hence g is not grounded. On the other hand, if g is not grounded, then there is some set k, such that $g \in k$ and all members of k are grounded (since g is the set of *all* the grounded sets). But then it follows that g would have to be grounded. So g is a paradoxical set - it both is and isn't grounded. This is quite a difficult problem to prove, especially in A3 where it took 345s of CPU time for Vampire to prove. The paradox does not occur in FDE$^{\to}$.

5 Conclusion and Future Work

This research has extended Belnap's "Useful Four-Valued Logic" with conditional connectives, and an automated theorem prover for the resulting logic has been implemented by translation to classical first-order logic. The addition of a conditional connective starts to make FDE truly computationally useful.

Future work on the theoretical side includes examining ways to overcome the functional incompleteness of FDE, e.g., by adding constants for the four truth values. On the practical side, further study on the use of FDE$^{\to_{cmi}}$ and FDE$^{\to_{con}}$ might reveal which is the most useful in applications, and a data driven framework for creating automated reasoning systems for these kinds of logics is planned.

#	Axioms	$\models$	Conjecture	RM3	A3	$FDE^{\rightarrow_{cmi}}$	$FDE^{\rightarrow_{con}}$
1		$\models$	$p \vee \neg p$	Yes	Yes	No	No
2	q	$\models$	$p \rightarrow q$	No	Yes	Yes	Yes
3	$\neg p$	$\models$	$p \rightarrow q$	No	No	No	No
4	$\neg(p \rightarrow q)$	$\models$	p	Yes	Yes	Yes	No
5		$\models$	$p \rightarrow (q \vee \neg q)$	No	Yes	No	No
6		$\models$	$p \rightarrow (p \vee \neg p)$	Yes	Yes	Yes	Yes
7		$\models$	$(p \wedge \neg p) \rightarrow q$	No	No	No	No
8	$p \vee q, \neg p$	$\models$	q	No	No	No	No
9		$\models$	$(\neg p \vee q) \leftrightarrow (p \rightarrow q)$	No	No	No	No
10		$\models$	$(p \leftrightarrow q) \leftrightarrow ((p \wedge q) \vee (\neg p \wedge \neg q))$	No	No	No	No
11		$\models$	$((p \rightarrow q) \wedge (q \rightarrow p)) \rightarrow (p \vee q \vee \neg(p \rightarrow q) \vee$ $\neg(q \rightarrow p) \vee ((\neg p \rightarrow \neg q) \wedge (\neg q \rightarrow \neg p)))$	Yes	Yes	No	Yes
12	$H(a)$	$\models$	$\exists x\, G(x) \rightarrow H(a)$	No	Yes	Yes	Yes
13		$\models$	$\exists x\, (G(x) \wedge \neg G(x)) \rightarrow H(b)$	No	No	No	No
14	$\exists x\, (G(x) \vee H(x)), \neg \exists y\, G(y)$	$\models$	$\exists z\, H(z)$	No	No	No	No
15	$H(a)$	$\models$	$\forall x\, (H(x) \rightarrow G(x)) \leftrightarrow$ $\forall x\, ((H(x) \wedge G(x)) \vee (\neg H(x) \wedge G(a)))$	No	No	No	No
16		$\models$	$\neg \exists y \forall x\, (E(x,y) \leftrightarrow \neg E(x,x))$	Yes	Yes	No	No
17		$\models$	$\forall z \exists y \forall x\, (E(x,y) \leftrightarrow (E(x,z) \wedge \neg E(x,x)))$ $\rightarrow \neg \exists w \forall u\, E(u,w)$	No	No	No	No
18		$\models$	$\neg \exists y \forall x\, (E(x,y) \leftrightarrow$ $\neg \exists z\, (E(x,z) \wedge E(z,x))$	Yes	Yes	No	No
19		$\models$	$\exists y \forall x\, (E(x,y) \leftrightarrow E(x,x)) \rightarrow$ $\neg \forall x \exists y \forall z\, (E(z,y) \leftrightarrow \neg E(z,x))$	Yes	Yes	No	No
20	$\forall y \exists z \forall x\, (E(x,z) \leftrightarrow x = y)$	$\models$	$\neg \exists w \forall x\, (E(x,w) \leftrightarrow \forall u\, (E(x,u) \rightarrow$ $\exists y\, (E(y,u) \wedge \neg \exists z\, (E(z,u) \wedge E(z,y)))))$	Yes	Yes	No	No

Table 6: Example Axiom-Conjecture pairs and their provability status

References

Arieli, O., and Avron, A. 1998. The Value of the Four Values. *Artificial Intelligence* 102(1):97–141.

Avron, A. 1991. Natural 3-Valued Logics: Characterization and Proof Theory. *Journal of Symbolic Logic* 56(1):276–294.

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

Belnap, N. 1977b. How a Computer Should Think. In Ryle, G., ed., *Contemporary Aspects of Philosophy*. Oriel Press. 30–56.

Belnap, N. 1992. A Useful Four-valued Logic: How a Computer Should Think. In Anderson, A.; Belnap, N.; and Dunn, J., eds., *Entailment: The Logic of Relevance and Necessity, Volume II*. Princeton UP. 506–541.

Dunn, J. 1976. Intuitive Semantics for First Degree Entailment and Coupled Trees. *Philosophical Studies* 29:149–168.

Hazen, A., and Pelletier, F. 2017. K3, Ł3, LP, RM3, A3, FDE, M: How to Make Many-Valued Logics Work for You. arXiv:1711.05816.

Kalish, D., and Montague, R. 1964. *Logic: Techniques of Formal Reasoning*. World, Harcourt and Brace.

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.

Łukasiewicz, J., and Tarski, A. 1930. Investigations into the Sentential Calculus. *Comptes Rendus des seances de la Societe des Sciences et des Lettres de Varsovie* 23:30–50.

Łukasiewicz, J. 1920. On Three-Valued Logic. *Ruch Filozoficny* 5:170–171.

Pelletier, F., and Urquhart, A. 2003. Synonymous Logics. *Journal of Philosophical Logic* 32:259–285. See also the authors' "Synonymous Logics: A Correction", *JPL 37*:95-100, 2008.

Pelletier, F.; Sutcliffe, G.; and Hazen, A. 2017. Automated Reasoning for the Dialetheic Logic RM3. In Rus, V., and Markov, Z., eds., *Proceedings of the 30th International FLAIRS Conference*, 110–115.

Priest, G. 1979. The Logic of Paradox. *Journal of Philosophical Logic* 8:219–241.

Rodrigues, O., and Russo, A. 1998. A Translation Method for Belnap Logic. Technical Report DoC 98/7, Imperial College London.

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