DOV GABBAY AND ANTHONY HUNTER

NEGATION AND CONTRADICTION

1 INTRODUCTION

The notion of falsity, denoted $\perp$, is fundamental in classical logic. If we have some set of assumptions, and we use classical proof rules to derive $\perp$, then there is some conflict in the assumptions. Let α and β denote two formulae, then the formula $\alpha \rightarrow (\beta \rightarrow \perp)$ expresses the fact that α and β are in conflict.

In classical logic, we can also represent conflict by using a negation symbol. So continuing the above example, we can represent the conflict between α and β by the following statement, where $\vdash$ denotes the classical consequence relation.

$$\{\alpha\} \vdash \neg\beta$$

In this way, negation and falsity are inter-changeable in classical logic. We can view $\perp$ as symbolising contradiction. We can read $\vdash (\alpha \rightarrow (\beta \rightarrow \perp))$ as α and β are contradictory — they are not acceptable together. This approach is independent of whether $\perp \vdash \alpha$ for an arbitrary α. $\perp$ is viewed as an atom which should not be derivable. It is unwanted but it may be consistent in some logics.

To support our discussion, we assume $\perp$ means 'contradiction'. We also assume that $\vdash (\alpha \wedge \beta) \rightarrow \perp$ means α and β are in 'conflict'. When $\perp \vdash \alpha$ holds for all α, then $\perp$ is 'falsity', 'conflict' means 'inconsistency'.

In this paper, we explore the relationship between negation and contradiction, in order to develop better techniques for handling inconsistent information. Intellectual activities usually involve reasoning with different perspectives. For example, consider negotiation, learning, or merging multiple opinions. Central to reasoning with different perspectives is the issue of handling conflict and hence inconsistencies. Yet our language for representing and reasoning with conflicts is underdeveloped. We lack some simple concepts to describe the nature of a logical conflict. For example, suppose agent 1 states that α holds, and agent 2 states that $\neg\alpha$ holds. What do we mean by conflict? What is the role of negation? Where does the inconsistency reside? What reasoning can either agent conduct with the opposing agent's statement?

A particular problem is that of the granularity of classical negation. For reasoning about contradictory information, classical negation, and hence classical inconsistency, is too general. Suppose we have three sets of propositions Δ, Γ and Φ, and suppose Δ is classically inconsistent with Φ, and Γ is classically inconsistent with Φ. We would like to know the 'degree of inconsistency' in each case. So does Δ contradict Φ 'more than' Γ contradicts Φ? Consider the following example.

EXAMPLE 1 Let Δ, Γ and Φ be defined as follows, where α, β, γ and δ are atoms.

$$\Delta = \{\neg\alpha, \neg\beta, \neg\gamma, \neg\delta\}$$

$$\Gamma = \{\neg\alpha, \beta, \gamma, \delta\}$$

$$\Phi = \{\alpha, \beta, \gamma, \delta\}.$$

Here, there is a clear criterion for claiming that Δ contradicts with Φ more than Γ contradicts with Φ: Since Δ has the complement of every literal in Φ, whereas Γ only has the complement of one literal in Φ. Whilst this is syntax sensitive, there are applications where we can attach equal significance to each of a set of literals.

For example, suppose A, B, and C are voters, each with equal significance, and α means A votes 'yes', $\neg\alpha$ means A votes 'no', and similarly β means B votes 'yes', and so on. Then Δ, Γ, and Φ represent different voting outcomes.

We attempt to address these questions in this paper. In the next section we formalize this notion of granularity by introducing notions of graded negation and graded toleration. In subsequent sections, we develop the notion and consider it in the context of reasoning about conflicts in information and conflicts between different viewpoints. We believe that negation is a conceptual building block that allows us to understand more about conflicts in information.

2 GRADED NEGATION AND GRADED TOLERATION

We extend the classical propositional language with connectives for graded negation and graded toleration, and extend the classical proof theory accordingly.

DEFINITION 2 *Let $\mathcal{C}$ be the usual set of classical formulae. The language $\mathcal{L}$ contains $\mathcal{C}$. Furthermore, if $\alpha, \beta \in \mathcal{L}$, then $\neg_\alpha\beta$ is in $\mathcal{L}$, and $\circ_\alpha\beta$ is in $\mathcal{L}$. The notation $\neg_\alpha$ is called graded negation. The notation $\circ_\alpha$ is called graded toleration. By $\neg_\alpha\beta$, we mean 'α negates β' and by $\circ_\alpha\beta$, we mean 'α tolerates β'.*

DEFINITION 3 *Let $\Delta \in \wp(\mathcal{L})$, and $\alpha, \beta, \neg_\alpha\beta \in \mathcal{L}$. Let $\vdash$ be classical consequence relation which we extend as follows,*

$$\Delta \vdash \neg_\alpha\beta \textit{ iff } \Delta \vdash \alpha \textit{ and } \{\alpha \wedge \beta\} \vdash \bot$$

Intuitively, for $\Delta \vdash \neg_\alpha\beta$, the formula $\neg_\alpha\beta$ captures the information that the inference α from Δ is in contention with β. This provides a succinct representation of the degree that Δ negates, or contradicts, β. Graded negation says more than classical negation since $\Delta \vdash \neg_\alpha \beta$ implies $\Delta \vdash \alpha \wedge \neg\beta$. As a corollary of the definition of graded negation, for any $\alpha \in \mathcal{L}$, $\vdash \neg_\alpha \neg\alpha \leftrightarrow \alpha$ and $\vdash \neg_{\neg\alpha} \alpha \leftrightarrow \neg\alpha$ hold. Also, $\vdash \neg_\top \bot$ and $\bot \vdash \neg_\bot \top$ hold.

DEFINITION 4 *Let $\Delta \in \wp(\mathcal{L})$, and $\alpha, \beta, \neg_\alpha\beta \in \mathcal{L}$. Let $\vdash$ be classical consequence relation which we extend as follows,*

$$\Delta \vdash \circ_\alpha \beta \textit{ iff } \Delta \vdash \alpha \textit{ and } \{\alpha \wedge \beta\} \nvdash \bot$$

Intuitively, for $\Delta \vdash \circ_\alpha\beta$, the formula $\circ_\alpha\beta$ captures the information that the inference α from Δ is consistent with β. Interesting axioms that we can derive from the definitions include the following.

$$\frac{\Delta \vdash \neg_\alpha\beta \wedge \neg_\alpha\gamma}{\Delta \vdash \neg_\alpha(\beta \wedge \gamma)} \qquad \frac{\Delta \vdash \circ_\alpha(\beta \wedge \gamma)}{\Delta \vdash \circ_\alpha\beta \wedge \circ_\alpha\gamma}$$

$$\frac{\Delta \vdash \neg_\alpha\gamma \wedge \neg_\beta\gamma}{\Delta \vdash \neg_{(\alpha\wedge\beta)}\gamma} \qquad \frac{\Delta \vdash \circ_{(\alpha\wedge\beta)}\gamma}{\Delta \vdash \circ_\alpha\gamma \wedge \circ_\beta\gamma}$$

$$\frac{\Delta \vdash \neg_{(\alpha\vee\beta)}\gamma}{\Delta \vdash \neg_\alpha\gamma \vee \neg_\beta\gamma} \qquad \frac{\Delta \vdash \circ_\alpha\gamma \vee \circ_\beta\gamma}{\Delta \vdash \circ_{(\alpha\vee\beta)}\gamma}$$

$$\frac{\Delta \vdash ((\circ_\alpha\beta \vee \circ_\alpha\gamma) \wedge \neg_\alpha\beta)}{\Delta \vdash \circ_\alpha\gamma}$$

Though the following do not normally hold.

$$\Delta \vdash \neg_\alpha\beta \vee \neg_{\neg\alpha}\beta$$

$$\frac{\Delta \vdash \circ_{\alpha\vee\beta}\gamma}{\Delta \vdash \circ_\alpha\gamma \vee \circ_\beta\gamma} \qquad \frac{\Delta \vdash \circ_\alpha(\beta \vee \gamma)}{\Delta \vdash \circ_\alpha\beta \vee \circ_\alpha\gamma}$$

$$\frac{\Delta \vdash \neg_\alpha\gamma \vee \neg_\beta\gamma}{\Delta \vdash \neg_{(\alpha\vee\beta)}\gamma} \qquad \frac{\Delta \vdash \neg_\alpha\beta \vee \neg_\alpha\gamma}{\Delta \vdash \neg_\alpha(\alpha \vee \gamma)}$$

For convenience, we introduce the following notation for the conjunction of formulae in a database.

DEFINITION 5 *For $\Delta \in \wp(\mathcal{L})$, if $\Delta = \{\phi_1, .., \phi_n\}$, then the symbol Δ is an abbreviation for $\phi_1 \wedge ... \wedge \phi_n$.*

There is a symmetry in graded negation, and similarly graded toleration, between pairs of mutually inconsistent databases as follows.

$$\Delta \vdash \neg_\alpha\Gamma \text{ iff } \Gamma \vdash \neg_{\neg\alpha}\Delta$$

$$\Delta \vdash \circ_\alpha\Gamma \text{ iff } \Gamma \vdash \circ_{\neg\alpha}\Delta$$

We now consider the negation of graded negation and graded toleration.

DEFINITION 6 *Let* $\Delta \in \wp(\mathcal{L})$*, and* $\alpha, \beta, \neg\neg_\alpha\beta, \neg\circ_\alpha\beta, \in \mathcal{L}$*. Let* $\vdash$ *be classical consequence relation which we extend as follows,*

$$\Delta \vdash \neg\neg_\alpha\beta \textit{ iff } \Delta \vdash \neg\alpha \textit{ or } \{\alpha \wedge \beta\} \not\vdash \bot$$

$$\Delta \vdash \neg\circ_\alpha\beta \textit{ iff } \Delta \vdash \neg\alpha \textit{ or } \{\alpha \wedge \beta\} \vdash \bot$$

From this definition, we can obtain axioms such as the following.

$$\frac{\Delta \vdash \neg_\alpha\beta}{\Delta \vdash \neg\circ_\alpha\beta} \qquad \frac{\Delta \vdash \circ_\alpha\beta}{\Delta \vdash \neg\neg_\alpha\beta}$$

$$\neg_\alpha\beta \vee \neg\neg_\alpha\beta$$

So far we have considered graded negation and graded toleration on the right-hand side of the consequence relation. However, we wish to also use it on the left-hand side. The way we do this is to assume that for a set of formulae on the left-hand side of the consequence relation, the following hold.

$\circ_\alpha\beta \in \Delta$ is an abbreviation for $\Delta \vdash \alpha$ and $\alpha \wedge \beta \nvdash \bot$
$\neg_\alpha\beta \in \Delta$ is an abbreviation for $\Delta \vdash \alpha$ and $\alpha \wedge \beta \vdash \bot$.

These assumptions mean that some formulae, such as $\circ_\alpha\neg\alpha$ can never be on the left-hand side.

The language of graded negation, and graded toleration, can be used to represent, and reason with, conflicts in information as illustrated by the following examples.

EXAMPLE 7 Suppose we want to capture statements such as 'If γ holds, then α negates β' and 'If α negates β, then it must tolerate γ'. We can represent these as follows.

$$\gamma \rightarrow \neg_\alpha\beta$$
$$\neg_\alpha\beta \rightarrow \circ_\alpha\gamma$$

EXAMPLE 8 Now suppose we represent 'α tolerates β or α tolerates γ' by $\circ_\alpha\beta \vee \circ_\alpha\gamma$, and '$\alpha$ negates β' by $\neg_\alpha\beta$. By using the definitions for graded negation and graded toleration, we can derive useful inferences such as $\circ_\alpha\gamma$ from these two formulae.

Though the consequence relation remains reflexive, supraclassical, transitive and monotonic, extending the classical consequence relation with graded negation and graded toleration is more than a conservative extension of classical logic: The extension incorporates a form of satisfiability checking into the object-level. This has ramifications on the computational properties of the logic.

3 MINIMUM NEGATION AND MAXIMAL TOLERATION

We can find a minimum α for $\neg_\alpha \delta$ for a given database. Similarly, we can find a maximal α for $\circ_\alpha \delta$ for a given database. Finding these can be useful, as we show in section 4, for discussing the nature conflicts between sets of formulae. For this, we extend the language with minimum negation and maximal toleration, as follows.

DEFINITION 9 *If $\alpha, \beta \in \mathcal{L}$, then $\ominus_\alpha \beta$ and $\oplus_\alpha \beta$ are in $\mathcal{L}$. The notation $\ominus_\alpha$ is called minimum negation, and $\oplus_\alpha$ is called maximal toleration. We say that $\ominus_\alpha \beta$ is 'α minimally negates β', and that $\oplus_\alpha \beta$ is 'α maximally tolerates β'.*

DEFINITION 10 *For $\Delta \in \wp(\mathcal{L})$, and $\neg_\alpha \delta, \neg_\beta \delta, \circ_\alpha \delta, \circ_\beta \delta \in \mathcal{L}$, the $\vdash$ relation is extended as follows.*

$$\Delta \vdash \ominus_\alpha \delta \textit{ iff } [\Delta \vdash \neg_\alpha \delta \textit{ and } \forall \beta \; [\textit{ if } \Delta \vdash \neg_\beta \delta \textit{ and } \vdash \alpha \rightarrow \beta \textit{ then } \vdash \beta \rightarrow \alpha]]$$

$$\Delta \vdash \oplus_\alpha \delta \textit{ iff } [\Delta \vdash \circ_\alpha \delta \textit{ and } \forall \beta \; [\textit{ if } \Delta \vdash \circ_\beta \delta \textit{ and } \vdash \beta \rightarrow \alpha \textit{ then } \vdash \alpha \rightarrow \beta]]$$

EXAMPLE 11 For $\Delta = \{\alpha \wedge \beta \wedge \gamma\}$, we obtain $\Delta \vdash \neg_\alpha \neg(\alpha \vee \beta)$, $\Delta \vdash \neg_{\alpha \wedge \beta} \neg(\alpha \vee \beta)$, $\Delta \vdash \neg_{\alpha \wedge \beta \wedge \gamma} \neg(\alpha \vee \beta)$, $\Delta \vdash \ominus_{(\alpha \vee \beta)} \neg(\alpha \vee \beta)$, and $\Delta \vdash \oplus_\gamma \neg(\alpha \vee \beta)$.

Axioms for minimum negation and maximal toleration include the following.

$$\frac{\Delta \vdash \ominus_\alpha \gamma \wedge \ominus_\beta \gamma}{\Delta \vdash \ominus_\alpha \gamma}$$

$$\frac{\Delta \vdash \oplus_\alpha \gamma \vee \oplus_\beta \gamma}{\Delta \vdash \circ_{\alpha \vee \beta} \gamma}$$

But axioms such as the following do not hold in general.

$$\frac{\Delta \vdash (\ominus_\alpha \beta \wedge \ominus_\alpha \gamma)}{\Delta \vdash \ominus_\alpha (\beta \wedge \gamma)}$$

$$\frac{\Delta \vdash \oplus_\alpha \gamma \vee \oplus_\beta \gamma}{\Delta \vdash \oplus_{(\alpha \vee \beta)} \gamma}$$

$$\frac{\Delta \vdash \oplus_\alpha \gamma \wedge \oplus_\beta \gamma}{\Delta \vdash \oplus_{(\alpha \wedge \beta)} \gamma}$$

The $\vdash$ consequence relation is supraclassical, reflexive, and transitive. However, symmetry does not hold in general for minimum negation or maximal toleration. So for example, for $\Delta \vdash \ominus_\alpha \Gamma$ and $\Gamma \vdash \ominus_\beta \Delta$, α is not usually equivalent to $\neg\beta$. Also, the consequence relation is non-monotonic, as illustrated by the following example.

EXAMPLE 12 Suppose $\Delta \vdash \ominus_\alpha \delta$, $\Delta \nvdash \neg_\beta \delta$ and $\alpha \vdash \beta$ hold: If $\Delta \cup \Gamma \vdash \neg_\beta \delta$ holds, then $\Delta \nvdash \ominus_\alpha \delta$.

We use the term 'minimum negation' because of the following result: If $\Delta \vdash \ominus_\alpha \gamma$ and $\Delta \vdash \ominus_\beta \gamma$ hold, then $\vdash \alpha \leftrightarrow \beta$ holds. However, we don't get the same result for toleration. For example, consider $\Gamma = \{\neg\phi \vee \neg\psi\}$, and $\Delta = \{\phi \wedge \psi\}$. Here, $\Delta \vdash \oplus_\phi \Gamma$ and $\Delta \vdash \oplus_\psi \Gamma$ hold.

4 CONFLICTING VIEWPOINTS

A motivation, that we offered in the introduction, for positing these new forms of negation was to address issues of handling conflicts between different perspectives. Let us call any classically consistent subset of $\mathcal{C}$ a viewpoint. We regard a viewpoint as a logical representation of a perspective. In this section we consider the situation when the union of two, or more, viewpoints are classically inconsistent.

Suppose Δ and Γ are viewpoints, where for some β, $\Delta \vdash \beta$ and $\Gamma \vdash \neg\beta$ hold. For this situation, what do we mean by negation? Where does the inconsistency reside? What reasoning can either agent conduct with the opposing agent's statement? Clearly, viewpoint Δ states the negation of viewpoint Γ's statement, and that this negation is symmetrical. Furthermore, for $\Delta \vdash \ominus_\alpha \Gamma$, the formula α indicates the source of the problematical data in Δ. Note, if $\Delta \cup \Gamma$ were consistent, then there would be no α such that $\Delta \vdash \ominus_\alpha \Gamma$.

Using minimal negation, we can represent stages in the resolution of conflicts between viewpoints. Suppose agent 1 has a viewpoint Δ, and agent 2 has a viewpoint Γ, it is quite likely that they cannot resolve their conflicts in one step. In other words, they cannot find a Δ^* and a Γ^* such that $\Delta^* \cup \Gamma^* \nvdash \bot$. This means the agents are expecting to find a series $\Delta_1, ..., \Delta_n$ and $\Gamma_1, ..., \Gamma_n$, where the conflict between them decreases. As a result there is a sequence of inferences, where Δ_1 is Δ, Γ_1 is Γ, and $\Delta_{n+1} \cup \Gamma_{n+1} \nvdash \bot$.

$$
\begin{array}{ll}
\Delta_1 \vdash \ominus_{\alpha_1} \Gamma_1 & \Gamma_1 \vdash \ominus_{\beta_1} \Delta_1 \\
\vdots & \vdots \\
\Delta_n \vdash \ominus_{\alpha_n} \Gamma_n & \Gamma_n \vdash \ominus_{\beta_n} \Delta_n
\end{array}
$$

Another aspect of this incremental resolution of conflicts is that the agents might wish to keep the contents of their viewpoints secret. This might be so they can get the best compromise out of the conflict resolution. In this way, the only information each agent has about the agent's viewpoint is that given by the minimal negation statements. In this circumstance, the agents would need to adopt some strategy such as agreeing at each stage i, the formulae α_{i+1} and β_{i+1} that are used in the next minimal negation statements.

5 CONFLICT ORDERING

To represent the relative weakenings of viewpoints, we can use the following definition of a conflict ordering.

DEFINITION 13 *For $\Delta, \Delta^*, \Gamma \in \wp(\mathcal{L})$, $\alpha, \beta \in \mathcal{L}$, the conflict ordering, denoted $\geq_\Gamma$, is defined as follows.*

$$\Delta \geq_\Gamma \Delta^* \textit{ if } [\textit{if } \Gamma \vdash \ominus_\alpha \Delta \textit{ and } \Gamma \vdash \ominus_\beta \Delta^* \textit{ then } \beta \vdash \alpha]$$

$$\Delta \geq_\Gamma \Delta^* \textit{ if } \Delta^* \cup \Gamma \nvdash \bot$$

We explain the conflict ordering as follows. Let Φ be a set of formulae, and let ψ be a formula such that (1) they are mutually inconsistent, i.e., $\Phi \cup \{\psi\} \vdash \bot$, and (2) there is no formula weaker than ψ that is inconsistent with Φ (i.e. there is no formula τ such that $\psi \vdash \tau$ and $\tau \nvdash \psi$ and $\Phi \cup \{\tau\} \vdash \bot$). Suppose we take an inferentially weaker set of formulae Φ^*, so that $\Phi \vdash \Phi^*$, and $\Phi^* \nvdash \Phi$. To maintain inconsistency, we cannot take a weaker formula than ψ, and indeed we may have to take a stronger formula ψ^*, so that $\psi^* \vdash \psi$ and $\psi \nvdash \psi^*$. In this way, the higher a set Δ_i is in the ordering $\geq_\Gamma$, the more Δ_i conflicts with Γ. This is illustrated by the following example.

EXAMPLE 14 Consider the following databases.

$$\begin{aligned} \Gamma &= \{\neg\alpha \wedge \neg\beta\} \\ \Delta &= \{\alpha, \beta\} \\ \Delta_1 &= \{\neg\alpha, \beta\} \\ \Delta_2 &= \{\beta\} \\ \Delta_3 &= \{\alpha, \beta, \gamma\} \end{aligned}$$

From this we obtain the following minimum negations and maximal tolerations for Γ.

$$\begin{array}{ll} \Gamma \vdash \ominus_{\neg\alpha\vee\neg\beta}\Delta & \Gamma \vdash \oplus_{\top}\Delta \\ \Gamma \vdash \ominus_{\neg\beta}\Delta_1 & \Gamma \vdash \oplus_{\neg\alpha}\Delta_1 \\ \Gamma \vdash \ominus_{\neg\beta}\Delta_2 & \Gamma \vdash \oplus_{\top}\Delta_2 \\ \Gamma \vdash \ominus_{\neg\alpha\vee\neg\beta}\Delta_3 & \Gamma \vdash \oplus_{\gamma}\Delta_3 \end{array}$$

Hence, $\Delta = \Delta_3 \geq_\Gamma \Delta_1 = \Delta_2$.

In the above example, we can see that comparing the inference $\Gamma \vdash \ominus_{\neg\alpha\vee\neg\beta}\Delta$ with the inference $\Gamma \vdash \ominus_{\neg\beta}\Delta_2$, the strength of the minimum negation increases from $\neg\alpha \vee \neg\beta$ to $\neg\beta$, as the set of formulae Δ is weakened to Δ_2. In this way, the minimal negation needs to draw a stronger inference from Γ in order to contradict with Δ_2.

Using the conflict ordering, we can identify weakenings of these viewpoints. So we can weaken Δ to Δ^*, where $\Delta >_\Gamma \Delta^*$, or we can weaken Γ to Γ^*, where $\Gamma >_\Delta \Gamma^*$, or weaken both. This then decreases, or even eliminates, the inferences that are in contention.

EXAMPLE 15 Consider the following sets.

$$\begin{aligned}\Delta_1 &= \{\neg\alpha \wedge \neg\beta \wedge \neg\gamma\}\\ \Delta_2 &= \{\neg\beta \wedge \neg\gamma\}\\ \Delta_3 &= \{\neg\gamma\}\\ \Gamma_1 &= \{\alpha \wedge \beta \wedge \gamma\}\\ \Gamma_2 &= \{\alpha \wedge \beta\}\end{aligned}$$

For these sets, two conflicting orderings are the following.

$$\Delta_1 \geq_{\Gamma_1} \Delta_2 \geq_{\Gamma_1} \Delta_3$$

$$\Gamma_1 \geq_{\Delta_3} \Gamma_2$$

So Δ_1 could be weakened to Δ_2, and then Δ_3. This could then be followed by Γ_1 being weakened to Γ_2. The viewpoints Δ_3 and Γ_2 are not in conflict.

We can also handle multiple viewpoints in this approach. Consider $\Delta, \Gamma, \Phi \in \wp(\mathcal{L})$, such that $\Delta \cup \Gamma \cup \Phi \vdash \bot$, but $\Delta \cup \Phi \not\vdash \bot$, $\Delta \cup \Gamma \not\vdash \bot$, and $\Gamma \cup \Phi \not\vdash \bot$. For this, we are interested in the following relationships, for some α, β and γ.

$$\Delta \vdash \ominus_\alpha(\Gamma \cup \Phi)$$

$$\Gamma \vdash \ominus_\beta(\Delta \cup \Phi)$$

$$\Delta \vdash \ominus_\gamma(\Delta \cup \Gamma)$$

EXAMPLE 16 Consider $\Delta = \{\alpha \vee \beta\}$, $\Gamma = \{\neg\alpha\}$, and $\Phi = \{\neg\beta\}$. Here, we get $\Delta \vdash \ominus_{\alpha\vee\beta}(\Gamma \cup \Phi)$, $\Gamma \vdash \ominus_{\neg\alpha}(\Delta \cup \Phi)$, and $\Phi \vdash \ominus_{\neg\beta}(\Delta \cup \Gamma)$. Furthermore, let $\Gamma^* = \{\neg\alpha \vee \alpha\}$. Hence we obtain $\Gamma \geq_{(\Delta\cup\Phi)} \Gamma^*$.

A conflict ordering $\geq_\Gamma$ is reflexive, transitive, and non-linear. Though it is not anti-symmetric. The minimum elements are the sets Δ_i such that $\Delta_i \cup \Gamma \not\vdash \bot$, and the maximum elements are the sets Δ_i such that $\Delta_i \vdash \bot$.

6 USING CONSISTENT SUBSETS

We can consider weakenings of viewpoints in terms of maximally consistent subsets of data and minimally inconsistent subsets of data, which we define below.

DEFINITION 17 *Let* $\Delta \in \wp(\mathcal{C})$, $Con(\Delta) = \{\Gamma \subseteq \Delta \mid \Gamma \not\vdash \bot\}$, *and* $Inc(\Delta) = \{\Gamma \subseteq \Delta \mid \Gamma \vdash \bot\}$.

$$MC(\Delta) = \{\Phi \in Con(\Delta) \mid \Psi \in Con(\Delta)\ \Phi \not\subset \Psi\}$$

$$MI(\Delta) = \{\Phi \in Inc(\Delta) \mid \Psi \in Inc(\Delta)\ \Psi \not\subset \Phi\}$$

We call $MI(\Delta)$ *the set of minimally inconsistent subsets of* Δ, *and* $MC(\Delta)$ *the set of maximally consistent subsets of* Δ.

If we weaken a viewpoint Δ to Δ^* and a viewpoint Γ to Γ^* so that $\Delta^* \cup \Gamma^* \not\vdash \bot$, it is not necessarily the case that $\Delta^* \cup \Gamma^*$ is a maximally consistent subset of $\Delta \cup \Gamma$. Furthermore, for $\Delta \vdash \ominus_\alpha \Gamma$, α is not necessarily a member of a minimally inconsistent subset of $\Delta \cup \Gamma$.

However, if we restrict Δ and Γ to containing only positive and negative literals, then we can identify a closer relationship. In this situation there is only one minimally inconsistent subset Φ of $\Delta \cup \Gamma$, and some maximally consistent subsets Ψ_1, ..,Ψ_n of $\Delta \cup \Gamma$. So $\Delta \cup \Gamma$ is $\Phi \cup \Psi_1 \cup ... \cup \Psi_n$. For $\Delta \vdash \ominus_\alpha \Gamma$, α is the disjunction of the literals in $\Delta \cap \Phi$. Similarly, for $\Gamma \vdash \ominus_\beta \Delta$, β is the conjunction of literals in $\Delta \cap \Psi_i$, for some maximally consistent subset Ψ_i.

EXAMPLE 18 Let $\Delta = \{\alpha, \beta\}$ and $\Gamma = \{\neg\beta, \gamma\}$. Hence, the maximally consistent subsets of $\Delta \cup \Gamma$ are $\Phi_1 = \{\alpha, \gamma, \beta\}$, and $\Phi_2 = \{\alpha, \gamma, \neg\beta\}$. The minimally inconsistent subset is $\{\beta, \neg\beta\}$. From this we obtain $\Delta \vdash \ominus_\beta \Gamma$, $\Delta \vdash \oplus_{\alpha\wedge\beta\wedge\gamma}\Gamma$, and $\Delta \vdash \oplus_{\alpha\wedge\gamma\wedge\neg\beta}\Gamma$.

In general the relationship between using consistent subsets of the databases to revise a viewpoint, and considering inferentially weaker viewpoints, is that the later is finer grained. For example, consider the viewpoint $\{\neg\alpha \vee \neg\beta, \beta\}$ which is inconsistent with α. If we wanted to eliminate the inconsistency, we could remove either $\neg\alpha \vee \neg\beta$ or β from the viewpoint. Alternatively, we could adopt a finer grained approach by for example weakening β to $\beta \vee \gamma$.

There are many open questions here. A particularly important one is what postulates (in the spirit of [10]) should we adopt for contracting viewpoints.

7 INCOHERENCE

In Sections 4 and 5, we considered conflicting pairs of databases or viewpoints. Now, we consider inconsistency within an individual database.

Some kinds of inconsistent database seem worse than others. For example, if someone is inconsistent on one topic, but not inconsistent on a number of other topics, then the inconsistency is localized, and they are in general coherent. Whereas, someone who is inconsistent on a number of topics is less coherent. Worse still is someone who is inconsistent on a number of inter-related topics.

Consider the database $\Delta_1 = \{\alpha \wedge \neg\beta, \alpha \wedge \beta, \neg\alpha \wedge \beta, \neg\alpha \wedge \neg\beta\}$. This database is totally incoherent, in the language made-up from the atoms $\{\alpha, \beta\}$, since any pair is inconsistent. A database is said to be n-incoherent if every n formulae from Δ is inconsistent. The database Δ_1 is 2-incoherent.

The focus of incoherence in a database is the minimally inconsistent subset of the database. The greater the proportion of the database that is in the minimally inconsistent subset, the greater the degree of incoherence.

Given an inconsistent database, it may be desirable to increase the coherence. To do this, we have choices. For example, consider $\Delta = \{\alpha, \gamma, \alpha \rightarrow \beta, \gamma \rightarrow \neg\beta, \gamma \rightarrow \delta, \alpha \rightarrow \epsilon\}$. This database is inconsistent. α and γ are in conflict. However, δ and ϵ follow from α and γ without the use of inconsistency. One can compromise by letting δ and ϵ remain in the database irrespective of what we take out to maintain consistency [7].

Another way of increasing coherence is to turn both $\alpha \rightarrow \beta$ and $\gamma \rightarrow \neg\beta$ into defeasible rules. Now we cannot deduce β nor $\neg\beta$ because neither has a clear undefeated proof. This approach — making the particpants defeasible — is due to Sanjay Modgil [13] — and is compatible with ideas about prioritizing formulae so as to resolve conflict. If we make a statement *more* defeasible according to how contradictory it is, we can use this information to decide which participants to ignore. Thus, if we extend Δ with $\{\alpha \rightarrow \psi, \gamma \rightarrow \neg\psi\}$, then α particpates in two proofs for a contradiction. This makes γ more preferred than α and δ more preferred than ϵ.

The ordering among inconsistencies may be externally imposed. Imagine an inconsistent database Γ is divided into several sections $\Gamma_1, \Gamma_2, \Gamma_3, ...$ though not necessarily disjoint. These maybe divided by topic, source, time, period, hypothetical or real world, or so on. Depending on how we split Γ, we can increase coherence in at least some subsets. An extreme case is to put all of the minimally inconsistent data into one exclusive section.

We now have to decide how to allow deductions from combinations of these sections. We may for example have a dominance ordering over them — so that if Γ_i dominates Γ_j, then formulae in Γ_i have to be included in Γ_j but not necessarily vice versa. To illustrate, Members of Parliament in the UK need to resign if they are personnally bankrupt. So an inconsistency in one database — their personal life — dominates another database — their political life.

Even though we can increase coherence by considering splitting a database into several sections, we may also need to have constraints on this division. We may need to force inconsistency in a particular section if there is an inconsistency in some combination of the other sections. For example, if a person is dishonest in their business life, then you might still consider that person honest with friends. However, if also know that person is unfaithful in their marriage, then you might wish to consider that they are inconsistent in other spheres of their life such as their friendships.

There are a number of choices for formalizing the way that formulae in one section can influence another. Alessandra Russo has studied some options where each section is repesented by a possible world in modal logic [15].

8 DISCUSSION

It is becoming more widely acknowledged that we need to develop more sophisticated means for handling inconsistent information. A better understanding of the notion of negation, and its relationship to contradiction, is important for this goal.

There have been a number of other approaches to addressing issues of inconsistency in data. First, there are the paraconsistent logics (for example [6; 1; 3]) that support weaker-than-classical non-trivial reasoning with inconsistent information. But these really only ignore inconsistency. They don't offer machinery for analysing the nature of the inconsistency.

Then there are truth maintenance systems [11; 8] and non-monotonic reasoning systems (for reviews see [4; 5] which assume some of the information is specified as defeasible, and the system then identifies plausible inferences on this basis. Again they don't offer machinery for analysing the inconsistency.

Modal logic, in particular epistemic logics (for review see [12]), are more closely related to the aim of providing a language for representing and reasoning with conflicts between viewpoints. However, they don't provide the language for capturing the 'degree of inconsistency' between viewpoints.

Also relevant are logics that reason with maximally consistent subsets of the data (for example [14; 2; 9]. These qualify inferences, so for example, inferences that follow from all maximally consistent subsets are preferred to inferences that only come from some maximally consistent subsets. However, these also don't provide machinery for analysing inconsistencies between viewpoints.

Finally there is belief revision theory [10]. This focusses on updating a database with a formula, where to maintain consistency, some of the database may have to be rejected. Hence, this doesn't address the needs of reasoning with conflicts, and in particular conflicts between viewpoints.

In conclusion, this paper offers ways in which we can develop machinery for handling different kinds of conflict in contradictory information. All this has been based on the classical notion of negation and contradiction.

ACKNOWLEDGEMENTS

This work was partially funded by the UK EPSRC as part of the VOILA project (GR J 15483).

Dov Gabbay
Imperial College, London

Anthony Hunter
University College, London

REFERENCES

[1] N. Belnap. A useful four valued logic. In *Modern Uses of Multiple-valued Logic*. G. Epstein, ed. Reidel, 1977.

[2] S. Benferhat, D. Dubois and H. Prade. How to infer from inconsistent beliefs without revising? In *Proceedings of the 14th International Joint Conference of Artificial Intelligence*. Morgan Kaufmann, 1995.

[3] Ph. Besnard and A. Hunter. Quasi-classical logic: Non-trivializable classical reasoning from inconsistent information. In *Symbolic and Quantitative Approaches to Uncertainty*, C. Froidevaux and J. Kohlas, eds. Lecture Notes in Computer Science. Springer, 1995.

[4] Ph. Besnard. *An Introduction to Default Logic*. Springer, 1989.

[5] G Brewka. *Common-sense Reasoning*. Cambridge University Press, 1991.

[6] N C da Costa. On the theory of inconsistent formal systems. *Notre Dame Journal of Formal Logic*, **15**, 497–510, 1974.

[7] F. Dargam. Compromised updates in labelled databases. In *Proceedings of the 3rd International Conference on Artificial Intelligence and Symbolic Mathematical Computations*, J. Calmet, J. Campbell and J. Pfalzgraf, eds. Lecture Notes in Computer Science. Springer, 1996.

[8] J. Doyle. A truth maintenance system. *Artificial Intelligence*, **12**, 231–272, 1979.

[9] M. Elvang-Goransson and A. Hunter. Argumentative logics: Reasoning from classically inconsistent information. *Data and Knowledge Engineering*, **16**, 125–145, 1995.

[10] P. Gärdenfors. *Knowledge in Flux*. MIT Press, 1988.

[11] J. De Kleer. An assumption-based TMS. *Artificial Intelligence*, **28**, 127–162, 1986.

[12] J. Meyer and W. van der Hoek. *Epistemic Logics for Artificial Intelligence and Computer Science*. Cambridge University Press, 1995.

[13] S. Modgil. Labelled deduction for non-monotonic reasoning. Technical report, Imperial College London, 1996.

[14] N. Rescher and R. Manor. On inference from inconsistent premises. *Theory and Decision*, **1**, 179–219, 1970.

[15] A. Russo. Generalising propositional modal logic using labelled deductive systems. In *FroCoS'96. Proceedings of the Ist International Workshop on Frontiers for Combining Systems*, Applied Logic, 1996. To appear.

GRAHAM PRIEST

WHAT NOT? A DEFENCE OF DIALETHEIC THEORY OF NEGATION

1 INTRODUCTION

The primary concern of logic is inference; and in particular, the question of what constitutes a valid inference. In investigating this issue, a certain class of notions has always appeared to be of crucial importance. We now call them logical constants, though they have been called by different names at different times. (For example, they were called *syncategoremata* by medieval logicians.) Much of logic has therefore been devoted to an analysis of these notions. Historically, the most contentious have been the quantifiers and the conditional. Consensus concerning the former has been achieved this century, due to the work of Frege and others. The debate concerning the latter shows no similar sign of convergence.

Amongst the logical constants, negation is, perhaps, the most crucial, dealing as it does with a certain polarity of thought, without which there could, some have thought, be no thought—or inference—at all. Historically, its behaviour may not have been terribly contentious. At least until this century. During this, our understanding of logical structures has become sharper and more profound by an order of magnitude that is historically unheard of; and this has allowed logicians to reflect on, and question, many traditional assumptions about the behaviour of negation. Two movements, in particular, stand out in this context: intuitionism and paraconsistency; the former can be seen as challenging the law of excluded middle; the latter as challenging the law of non-contradiction.

For these reasons, the nature of negation is a contemporary question that is both important and difficult. In this essay, I want to address it and suggest a dialetheic answer.[1]

2 NEGATION OR NEGATIONS?

How, then, does negation behave?[2] There is a short way with this question. There is no such thing as negation; there are lots of different negations: Boolean negation,

[1] Sainsbury [24, p. 142], discerns a challenge for dialetheism: to provide an account of what understanding negation involves. I hope that this essay goes a reasonable way towards meeting that challenge.

[2] I will concern myself only with propositional negation, though this fits into a much broader family of negative constructions. See Sylvan, [28].

intuitionist negation, De Morgan negation. Each of these behaves according to a set of rules (proof-theoretic or semantic); each is perfectly legitimate; and we are free to use whichever notion we wish, as long as we are clear about what we are doing. If this is right, there is nothing left to say about the question, except what justifies us in categorising a connective as in the negation family. And I doubt that there is anything very illuminating to be said about that. Virtually *every* negation-like property fails on some account of a connective that is recognisably negation-like: the law of excluded middle, the law of non-contradiction, double negation, De Morgan's laws, contraposition, and so on. All we are left with is a family-resemblance whose fluid boundaries are largely historically determined.

I do not think that the answer is right, however. It makes a nonsense of too many important debates in the foundations of logic. Doubtless, philosophical debates do rest on confusion sometimes, but questions concerning the role of negation in discourses on infinity, self-reference, time, existence, etc., are not to be set aside so lightly.

At the root of this kind of answer is a simple confusion between a theory and what it is a theory of.[3] We have many well worked-out theories of negation, each with its own proof-theory, model-theory and so on. And if you call the theoretical object constituted by each theory a negation, then, so be it: there are many negations. But this does not mean that one can deploy each of these theoretical objects at will and come out with the correct answer. The theoretical object has to fit the real object; and how this behaves is not a matter of choice.

A comparison with geometry may be helpful here. There are, in a sense, many geometries. Each has its own well defined structure; and, as an abstract mathematical structure, is worthy of investigation. But if we think of each geometry, not as an abstract mathematical structure, but, suitably interpreted, as a theory about the spatial (or spatio-temporal) structure of the cosmos, we are not free to choose at will. The theory must answer to the facts—or, if one is not a realist, at least cohere in the most satisfactory way with the rest of our theorising.

There is always an extreme conventionalist line to be run here. One might say, as Poincaré [12] did, that we are free to choose our geometry at will, e.g. on the grounds of simplicity, and then fix everything else around it. Similarly, we might insist that we are free to employ a certain notion of negation and make everything else fit. But such a line is not only philosophically contentious, but foolhardy, at least in advance of a good deal of further investigation. The tail may end up wagging a dog of a considerable size. For example, as Prior [20] pointed out a long time ago, we can determine to use a connective $*$ (tonk) according to the rules of

[3] See Priest [14, Ch. 14]. The confusion is manifested by, e.g. Quine [21, p. 81] when he complains that someone who denies *ex contradictione quodlibet* just doesn't know what they are talking about, since changing the laws is changing the subject. A similar confusion is apparent in those who argue that someone who suggests adopting a non-classical logic wants to revise logic, i.e. correct reasoning. Such a person need only be suggesting a revision of a theory of logic, *not* logic itself. One cannot simply *assume* that classical logic gets it right. That's exactly what is at issue here.

inference $\alpha \vdash \alpha * \beta$ and $\alpha * \beta \vdash \beta$. But the cost of this is accepting that if anything is true, everything is!

3 CONTRADICTORIES

We see, then, that a simple voluntarism with respect to negation is unsatisfactory. If it is to be applied, an account of negation must be considered not just as an abstract structure, but as a theory *of* something, just as a geometry is a theory of physical space. And this will put substantial constraints on what an acceptable account is.

The next question is what, exactly, an account of negation is a theory of. It is natural to suggest that negation is a theory of the way that the English particle 'not', and similar particles in other natural languages, behaves. This, however, is incorrect. For a start, 'not' has functions in English which do not concern negation. For example, it may be used to reject connotations of what is said, though not its truth, as in, for example, 'I am not his wife: he is my husband'.[4]

More importantly, negation may not be expressed by simply inserting 'not'. For example, the negation of 'Socrates was mortal' may be 'Socrates was not mortal'; but, as Aristotle pointed out (*De Interepretatione*, ch. 7), the negation of 'Some man is mortal' is not 'Some man is not mortal', but 'No man is mortal'.

These examples show that we have a grasp of negation that is independent of the way that 'not' functions, and can use this to determine when 'notting' negates. But what is it, then, of which we have a grasp? We see that there appears to be a relationship of a certain kind between pairs such as 'Socrates is mortal' and 'Socrates in not mortal'; and 'Some man is mortal' and 'No man is mortal'. The traditional way of expressing the relationship is that the pairs are *contradictories*, and so we may say that the relationship is that of contradiction. Theories of negation are theories about this relation.

As usual in theorisation, we may reach a state where we have to reassess the situation. For example, it may turn out that there are several distinct relationships here, which need to be distinguished. But at least this is the data to which theorisation must (and historically did) answer, at least initially.

Having got this far, the next obvious question is what the relationship of contradiction is a relationship between: sentences, propositions, some other kind of entity? There are profound issues here; but, as far as I can see, they do not affect the question of negation substantially. For any issue that arises given one reasonable answer to this question, an equivalent one arises for the others. So I shall simply call the sorts of thing in question, non-commitally, statements, and leave it at that.

[4] See, e.g. Horn [9, pp. 370 ff].

4 THE LAWS OF EXCLUDED MIDDLE AND NON-CONTRADICTION

So if α is any statement, let $\neg\alpha$ represent its contradictory. (Contradictories, unlike contraries and sub-contraries are unique—at least up to logical equivalence.) What relationships hold between these? Traditional logic and common sense are both very clear about the most important one: we must have at least one of the pair, but not both.[5] It is precisely this which distinguish contradictories from their near cousins, contraries and sub-contraries. If we have two contraries, e.g., 'Socrates was black' and 'Socrates was white', it is necessarily false that Socrates was black $\wedge$ Socrates was white; but it is not necessarily false that Socrates was black $\vee$ Socrates was white. Dually, if we have two subcontraries, e.g., 'Socrates was under 2m. tall' and 'Socrates was over 1m. tall', it is necessarily false that Socrates was under 2m. tall $\vee$ Socrates was over 1m. tall, but not necessarily true that Socrates was under 2m. tall $\wedge$ Socrates was over 1m. tall.

This fact about contradictories obviously gives immediately two of the traditional laws of negation, the law of excluded middle (LEM), $\alpha \vee \neg\alpha$, and the law of non-contradiction (LNC), $\neg(\alpha \wedge \neg\alpha)$.[6] (Note that the LNC, unlike the LEM, is not only a principle about contradictories, but is itself a negative thesis. This is important, and we will return to it later.) Now, maybe the traditional claim about contradictories—and consequently these two laws—is wrong; but it would certainly seem to be the default position. The onus of proof is therefore on those who would dispute it.

Disputation comes from at least two directions. The first is that of some (though not all) paraconsistent logicians. The argument here is that some contradictories are both true, i.e., for some βs we have $\beta \wedge \neg\beta$. We do not, therefore, have $\neg(\beta \wedge \neg\beta)$. We will look more closely at the first part of this argument later. For the moment, just note that if it is correct, it undercuts the second part of the argument (at least without some further considerations). For if some contradictions are true, we may well have both $\beta \wedge \neg\beta$ and $\neg(\beta \wedge \neg\beta)$. Hence, the fact that some contradictions are true does not, of itself, refute the LNC (at least in the form in question here).

The second direction from which one might dispute the traditional characterisation is that of some logicians who suppose there to be sentences that are neither true nor false, notably intuitionist logicians. The argument here is that if α is neither true nor false, so is $\neg\alpha$. Hence, assuming that disjunction behaves normally, $\alpha \vee \neg\alpha$ is not true.[7] The claim that certain statements are neither true nor false is clearly a substantial one. The claim that disjunction behaves normally is also challengeable. (If we give a supervaluationist account, $\alpha \vee \neg\alpha$ may be true even

---

[5] Classically, these facts actually characterise contradictories up to logical equivalence. This, however, is moot. If β satisfies the conditions $\Box\neg(\alpha \wedge \beta)$ and $\Box(\alpha \vee \beta)$, and γ is any necessary truth, then so does $\beta \wedge \gamma$; but β does not entail $\beta \wedge \gamma$ unless one identifes entailment with strict implication.

[6] I express the laws in the form of schemas. I will use lower case Greek letters schematically throughout this essay.

[7] If conjunction behaves normally, the LNC may also fail for truth-valueless sentences.

though each disjunct fails to be so.) However, we need discuss neither of these issues here. For from the present perspective there is an obvious objection. If $\neg$ behaves as suggested, it is not a contradictory-forming operator at all, but merely a contrary-forming one. This would seem particularly clear if we consider the intuitionist account of negation. According to this, $\neg\alpha$ is true (= assertable) just if there is a proof that there is no proof that α. This is obviously a *contrary* of α.[8]

A genuine contradictory-forming operator will be one that when applied to a sentence, α, covers *all* the cases in which α is not true. Thus, it is an operator, $\neg$, such that $\neg\alpha$ is true iff α is either false or neither true nor false. (In English, such an operator might be: it is not the case that.) For this notion, which is the real contradictory-forming operator, the LEM holds.

Those who believe in simple truth-value gaps would seem to have little reply to this objection. The intuitionist does have a reply to hand, however. They can argue that a contradictory-forming operator, as traditionally conceived, literally makes no sense.[9] The argument is a familiar one from the writings, notably, of Dummett.[10] *In nuce*, it is as follows. If a notion is meaningful there must be something that it is to grasp its meaning. Whatever that is, this must be manifestable in behaviour (or, the argument sometimes continues, the notion would not be learnable). But there is no suitable behaviour for manifesting a grasp of a connective satisfying the conditions of a classical contradictory-forming operator. In particular, we cannot identify the behaviour as that of being prepared to assert $\neg\alpha$ when (and only when) α fails to be true. For this state of affairs may well obtain when there is no principled way for us to be able to recognise that it does.

There are subtle issues (and a substantial body of literature) here. And to deal with them satisfactorily would require taking up a disproportionate part of this essay. But let me at least say something about the matter. For a start, I do not see why the grasp of a notion must be manifestable. There is no reason why, in general, certain notions should not be hard-wired in us. If, for example, there is a Fodor-style language of thought,[11] it is quite natural to suppose that single-bit toggling is a primitive operation. One might even tell an evolutionary story as to how this came about: it is the simplest and most efficient mechanism for implementing the polarity of thought. In particular, then, a contradictory operator does not have to be learned; its use is merely triggered in us by certain linguistic contexts, in much the

[8] Most perspicuously, consider the embedding of intuitionist logic into $S4$ where the modal operator $\Box$ is considered as a provability operator. Then $\neg\alpha$ is translated into $\Box\neg\alpha^+$ (where α^+ is the translation of α). In other words, $\neg\alpha$ is intuitionistically true iff the negation of α is *provable*.

[9] They might even suggest that $\neg T\langle\alpha\rangle \to T\langle\neg\alpha\rangle$ is perfectly acceptable provided the negation in the antecedent is understood as intuitionist negation. But this is highly problematic, for it leaves them no way of expressing their view concerning instances of the Law of Excluded Middle that fail: if α is undecided, one can no longer say that $\alpha \vee \neg\alpha$ is not true, let alone false, since $\neg T\langle\alpha \vee \neg\alpha\rangle$ now entails $T\langle\neg\alpha \wedge \neg\neg\alpha\rangle$.

[10] See, e.g. Dummett [4], esp. pp. 224–5 of the reprint. A somewhat different argument is explained and dispatched in Read [22, pp. 220–230].

[11] See Fodor [6].

same way that the categories of universal grammar are, according to Chomsky.[12]

But even granting that the grasp of a notion must be manifestable, I do not see why it must be manifestable by anything as strong as the argument requires (which is, I agree, impossible). In particular, it can be manifested by being prepared to assert $\neg\alpha$ when in a position to recognise that α fails to be true, and refusing to assert it when in a position to recognise that α is true.[13] It could well be suggested that such a manifestation would not be adequate. There will be many cases where we are not in a position to recognise either. People could therefore manifest the same behaviour whilst disagreeing about how to handle new cases when these become recognisable, and so meaning different things. This is true. But if the people not only behave as suggested, but also manifest a disposition to agree on new cases, this is sufficient to show (if not, perhaps, conclusively, then at least beyond reasonable doubt) that they are operating with the notion in the same way. In just this way, the fact that we are all prepared to apply, or refrain from applying, the word 'green' to hitherto unseen objects when they come to light, shows that we all mean the same thing by the word. This is essentially what following an appropriate rule comes to, in Wittgensteinian terms.[14]

There is much more to be said here. But if the onus of proof is on an intuitionist, as it would seem to be in the case of a contradictory-forming operator, I know of no argument against the LEM that I find persuasive. (That one can tell a coherent epistemological/metaphysical intuitionist story is not at issue.)

Before we leave the LEM it is worth noting that the fact that for every pair of contradictories one must be true (period), does not entail that for every situation one of each pair must be true *of it*. If one thinks of a situation as *part* of the world, then it may well be argued that neither of a pair of contradictories need be true of it. Thus, consider the situation concerning my bike. It may be the case that neither 'Gent is in Belgium' nor 'Gent is not in Belgium' is true of this situation. See Restall [23]. The question of whether or not one needs to consider partial situations, as so conceived, is important in discussions of the semantics of conditionals. But since conditionality is not the issue here, I will discuss the matter no further.

5 TRUTH AND FALSITY

So far, we have met two of the classical laws of negation, LEM and LNC. A third, the law of double negation (LDN) is simply derivable. The relationship of being contradictories is symmetric. That is, if β is the contradictory of α, then α is the contradictory of β. In particular, α is the contradictory of $\neg\alpha$. Hence, $\neg\neg\alpha$ just is α.

---

[12] See, e.g., Chomsky [3].

[13] A different suggestion, though not one I would make, is that an understanding can be manifested by using classical logic. This raises quite different issues.

[14] See *Philosophical Investigations*, Part I, esp. sections 201-40.

We are now in a position to look at another important feature of negation: its truth conditions. To do this we will need a definition of falsity. Let us define 'α is false' to mean that $\neg\alpha$ is true. This is not the only plausible definition; one might also define it to mean that α is not true. It may turn out that these two definitions are equivalent, of course. However, to assume so here would be to beg too many important questions. And the present definition is one that all parties can agree upon, classical, intuitionist and paraconsistent.

The definition of falsity assures us that $\neg\alpha$ is true iff α is false. Dually, $\neg\alpha$ is false iff $\neg\neg\alpha$ is true (by the definition of falsity) iff α is true, by LDN. Hence, the traditional understanding of the relationship between truth and falsity falls out of the understanding of negation as contradiction, and the definition of falsity.

Two more of the classical laws of negation, the Laws of De Morgan (LDM), can now also be dealt with. These involve conjunction and disjunction essentially; and so we need to make some assumption about how they behave. Since this is not an essay on conjunction/disjunction, this does not seem the place to discuss the matter at great length. For present purposes, let us suppose that they behave as tradition says they do: a conjunction is true iff both conjuncts are true, and false iff at least one conjunct is false. The conditions for disjunction are the obvious dual ones.

One of De Morgan's Laws is the equivalence of $\neg(\alpha \wedge \beta)$ and $\neg\alpha \vee \neg\beta$. This can now be demonstrated thus: $\neg(\alpha \wedge \beta)$ is true iff $\alpha \wedge \beta$ is false iff α is false or β is false iff $\neg\alpha$ is true or $\neg\beta$ is true iff $\neg\alpha \vee \neg\beta$ is true. Dually, $\neg(\alpha \wedge \beta)$ is false iff $\alpha \wedge \beta$ is true (LDN) iff α and β are true iff $\neg\alpha$ and $\neg\beta$ are false (LDN) iff $\neg\alpha \vee \neg\beta$ is false. The other of De Morgan's Laws is an equivalence between $\neg(\alpha \vee \beta)$ and $\neg\alpha \wedge \neg\beta$, and can be verified by a similar argument.

The connection between negation and the conditional is more difficult to deal with, but this is because the conditional is itself more contentious. Indeed, the claim that there are different kinds of conditional (entailments, causal conditionals, indicative conditionals, subjunctive conditionals) is well known; some of these distinctions are well motivated; and negation may well interact with different conditionals in different ways. A minimal condition for a conditional of any kind would seem to be that it preserve truth in an appropriate way from antecedent to consequent. From this, it follows that *modus ponens*, $\alpha, \alpha \rightarrow \beta \vdash \beta$, is valid. The most important question concerning a conditional in the present context is whether it preserves falsity in the reverse direction. For some conditionals, at least, this would seem to fail, as, e.g., Stalnaker and Lewis have argued.[15] And if $\alpha \rightarrow \beta$ fails to preserve falsity backwards, $\neg\beta \rightarrow \neg\alpha$ will fail to preserve truth forwards, and so will not be true. The law of contraposition (LC), $\alpha \rightarrow \beta \vdash \neg\beta \rightarrow \neg\alpha$ is not, therefore, to be expected to hold for an arbitrary conditional. Of course, there may well be conditionals which do preserve falsity in the appropriate way; in fact one can always *define* one, $\Rightarrow$, in a simple fashion: $\alpha \Rightarrow \beta$ is just $(\alpha \rightarrow \beta) \wedge (\neg\beta \rightarrow \neg\alpha)$. For such conditionals contraposition will hold.

[15] See, e.g., Stalnaker [27], Lewis [10]. See also Priest [14, 6.5].

In this section, I have talked of truth. I have said nothing about truth-in-an-interpretation, as required, for example, for a model-theoretic account of validity. It is important to distinguish these two notions, for they are often confused. The first is a property (or at least a monadic predicate); the second is a (set-theoretic) relation. It is natural enough to suppose that truth is at least coextensive with truth-in-g, where g is some one privileged interpretation (set). And this may provide a constraint on the notion of truth-in-an-interpretation. But it, even together with an account of truth, is hardly sufficient to determine a theory of truth-in-an-interpretation. It does not even determine, for example, how to conceptualise an interpretation. So how are an account of truth-in-an-interpretation, appropriate for the connectives we have been discussing, and a corresponding model-theoretic notion of validity, to be formulated? The details of this are a bit more technical than the rest of this essay, and I will defer them to an appendix: the rest of the material does not presuppose them.

6 TRUTH AND CONTRADICTION

Starting with a conception of negation as a contradictory-forming operator, we have now validated five standard laws of negation (LEM, LNC, LDN and the two LDM), and a sixth (LC) in certain contexts. We have hardly settled all the central issues concerning negation, however.

It is common to distinguish between the LEM and the Principle of Bivalence: every statement is either true or false.[16] Though these are natural mates, either can hold without the other, given the right account of other things. Similarly, we need to distinguish between the LNC and what I will call, for want of a better term, the Principle of Consistency: no statement is both true and false. Again, though these are natural mates, it is quite possible to have one without the other. In particular, as I have already observed, the fact that every instance of $\neg(\alpha \wedge \neg\alpha)$ is true does not, on its own, prevent some instances of α and $\neg\alpha$ from being true. So what is one to say about the Principle of Consistency? This is the next issue that needs to be addressed.

The traditional view endorses this Principle. But the traditional view has been called into question by some paraconsistent logicians, who assert that some contradictions are true, dialetheists. The case for this is a long one, and, like the intuitionist case against classical negation, is too long to take up in detail here; but let me say a little.[17]

Many examples of dialetheias have been suggested, but the most impressive ones are those generated by the paradoxes of self-reference. Here we have a set of arguments that appear to be sound, and yet which end in contradiction. *Prima facie*, then, they establish that some contradictions are true. Some of these arguments

[16] See Haack [8, p. 66f].
[17] The case is made in Priest [14].