n.k.jones@bham.ac.uk Draft of: 21 October 2018

The Proper Treatment of Identity in Dialetheic Metaphysics

Nicholas K. Jones

Abstract

According to one prominent tradition in mainstream logic and metaphysics, identity is indistinguishability. Priest has recently argued that this permits counterexamples to the transitivity and substitutivity of identity within dialetheic metaphysics, even in paradigmatically extensional contexts. This paper investigates two alternative regimentations of indistinguishability. Although classically equivalent to the standard regimentation on which Priest focuses, these alternatives are strictly stronger than it in dialetheic settings. Both regimentations are transitive, and one satisfies substitutivity. It is argued that both regimentations provide better candidates to occupy the core theoretical role of numerical identity than does the standard regimentation.

According to one prominent strand of mainstream logic and metaphysics, numerical identity satisfies each of the following:[1]

Substitutivity of Identicals (SI) $a = b, \phi_x(b) \vdash \phi_x(a)$

Transitivity of Identity (TI) $a = b, b = c \vdash a = c$.

Identity is Indistinguishability (II) Numerical identity is indistinguishability, understood as complete sharing of properties.

[1] Notation and terminology. $\rightarrow$ is the material conditional, with $\phi \rightarrow \psi$ definitionally equivalent to $\neg\phi \vee \psi$. $\leftrightarrow$ is the material biconditional, with $\phi \leftrightarrow \psi$ definitionally equivalent to $(\phi \rightarrow \psi) \wedge (\psi \rightarrow \phi)$. $\forall X$ is the second-order universal quantifier, with X a variable of the syntactic type of monadic predicates. ϕ and ψ are schematic letters for formulae. a, b, and c are used ambiguously, sometimes as constants and sometimes as schematic for any constant or variable of the syntactic type of singular terms. For each formula ϕ, variable v, and constant c of the same syntactic type as v, $\phi_v(c)$ is the result of replacing each free occurrence in ϕ of v with an occurrence of c; for example, $(x = c)_x(b)$ is $b = c$, and $Xa_X(F)$ is Fa. $\Gamma \vdash \phi$ says that closed formula ϕ is derivable from collection Γ of closed formulae. A consequence relation $\models$ validates $\Gamma \vdash \phi$ iff $\Gamma \models \phi$.

These theses are not unrelated. TI is an instance of SI where ϕ is $x = c$, and therefore follows from SI. Moreover, both SI and TI are valid in classical logic, if the standard second-order regimentation of indistinguishability as $\forall X(Xa \leftrightarrow Xb)$ is used to define the identity sign. Since II appears to license that definition of the identity sign, II appears to validate SI and TI.

Perhaps the most controversial of the above three principles is II. Doubts flow primarily from two sources.[2] One concerns putative examples of numerically distinct indistinguishable objects. The other concerns putative circularities arising from identifying numerical identity with, or grounding it in, indistinguishability. My own view is that this second kind of consideration is compelling, though I won't go into details here. My present claim is not that II is true, or unproblematic, or even close to universally accepted; I claim only that II, together with SI and TI, belongs to a prominent, reasonably mainstream, and not obviously false conception of identity. I propose simply to bracket doubts about II here, in the interest of exploring the view that results.

In a series of papers culminating in his recent book *One*, Graham Priest presents a powerful challenge to this view.[3] Central to this challenge is a dialetheic theory of identity on which II holds and yet both SI and TI are invalid. Although putative counterexamples to SI abound, notably involving attitude ascriptions and intensional operators, Priest's theory strikingly permits failures of SI and TI in even paradigmatically extensional contexts. He argues persuasively that failures of SI and TI allow for novel and attractive solutions to a cluster of seemingly intractable problems concerning the unity of propositions and other objects, intentionality, change, modality, vagueness, fission, and fusion.[4]

Attractive as these solutions may be, one might suspect this view of changing the subject: it uses the identity sign to express a non-transitive identity-like relation, rather than the transitive relation of identity properly so-called. That's where II comes in; it helps to alleviate this suspicion by ensuring that these rival views assign different logical properties to the same relation of indistinguishability, as defined in the standard second-order way, rather than merely differing over which relation they use to interpret the identity sign.

This paper investigates two alternative regimentations of the intuitive notion of complete sharing of properties. Although classically equivalent to the standard regimentation, they are strictly stronger than it in dialetheic settings. If the first regimentation captures numerical identity, then TI but not SI is valid. And if the second regimentation captures numerical identity, then both TI and SI are valid. I will argue that both regimentations provide better candidates to occupy the core theoretical role of numerical identity than does the standard regimentation. If these claims are correct, the non-transitive relation defined by $\forall X(Xa \leftrightarrow Xb)$ is not

[2] For discussion and citations, see, e.g., (Hawthorne, 2003, §§1–2), (Hawley, 2009), (Forrest, 2016), and (Noonan and Curtis, 2018, §§1–2).

[3] (Priest, 2009), (Priest, 2010b), (Priest, 2010a), (Priest, 2014, especially chapters 2 and 5).

[4] For discussion of Priest's solution to the problem of unity and two alternative consistent solutions to that problem, see (Jones, 2018).

numerical identity, and is therefore irrelevant to the identity-involving problems mentioned above.

Unlike previous dialetheic attempts to validate SI and TI in the presence of II, my approach requires no departure from Priest's preferred paraconsistent logic LP, or from attractive meta-rules like the transitivity of entailment.[5] The focus throughout is on what are normally regarded as extensional contexts, setting intensional operators and attitude ascriptions aside as raising complications not central to the present investigation; for Priest's system permits failures of SI and TI not involving intensionality or the attitudes.[6]

1 From indistinguishability to substitutivity

Our first task is to regiment the notion of indistinguishability, or complete sharing of properties, i.e.: every property of a is a property of b, and conversely. This quantification over properties is naturally regimented as second-order quantification. We can then define our first notion of indistinguishability thus:

a is weakly indistinguishable from $b =_{df} \forall X(Xa \leftrightarrow Xb)$.

I will assume for the time being that the notion of indistinguishability employed in II is weak indistinguishability. If II is true, the identity sign can be introduced into a second-order language without it, by defining $a = b$ as the second-order formula $\forall X(Xa \leftrightarrow Xb)$. This way of introducing the identity sign guarantees that it has its intended interpretation, because II says that numerical identity is weak indistinguishability. When the identity sign is defined in this way, SI and TI are valid in classical second-order logic.

Priest adopts a paraconsistent logic (LP) within which this second-order definition of identity does not validate SI or TI. He also provides a corresponding model-theoretic semantics for the language of second-order logic which supplies countermodels to, and thereby invalidates, both SI and TI when the identity sign is defined by weak indistinguishability. I'll describe some such countermodels in §2.4 and §2.5. Instead, this semantics validates only the following weaker substitutivity principles:

Material SI (MSI) $a = b \vdash \phi_x(b) \rightarrow \phi_x(a)$

[5] For these alternatives, see (Cobreros et al., 2014) and (Cobreros et al., 2013).

[6] The underlying mechanisms responsible for failures of SI and TI in Priest's system fundamentally differ between extensional and intensional contexts. In intensional contexts, SI and TI fail because many functions from worlds to extensions are not in the second-order domain. The specifically dialetheic aspect of the semantics is irrelevant, in that failures of SI and TI for intensional contexts remain even if all properties are classical (i.e. extension and anti-extension are exclusive as well as exhaustive at every world). However, if every property is classical and every formula defines a property, failures of SI and TI vanish. By contrast, if all properties are classical, Priest's semantics permits no failures of SI and TI in extensional contexts; but if properties are not all classical, failures of SI and TI for extensional contexts remain even if all formulae define properties. See §2.6 and (Priest, 2014, pp35–37) for details about some of these claims.

Material TI (MTI) $a = b \vdash b = c \rightarrow a = c$

MTI is an instance of MSI where ϕ is $x = c$, and therefore follows from MSI.

At a first pass, one might expect that SI and TI follow from MSI and MTI. For one might expect that one can use MSI to argue from the left of SI to its right, thus:

(1) Suppose: $a = b$

(2) Suppose: $\phi_x(b)$

(3) By (1), MSI: $\phi_x(b) \rightarrow \phi_x(a)$.

(4) By (2), (3), *modus ponens*: $\phi_x(a)$.

However, this argument fails because *modus ponens* is invalid in *One*'s paraconsistent system.[7]

The model theory adopted in *One* invalidates *modus ponens* because it permits models in which (i) ϕ is true, (ii) $\phi \rightarrow \psi$ is true, and yet (iii) ψ is false. The trick is to permit models that make some formulae both true and false, as well as to make $\phi \rightarrow \psi$ true in a model whenever ϕ is false in that model *including whenever* ϕ *is both true and false in the model.* When ϕ is both true and false, (i) holds because ϕ is true, and (ii) holds because ϕ is false. But (iii) may also hold because merely making ϕ true and false is irrelevant (in this semantics) to whether any other, arbitrarily chosen and potentially unrelated formula ψ is true. §2.2 onwards discusses this semantics in more detail.

We've seen where the arguments from MSI and MTI to SI and TI break down in Priest's setting. Let's now see why MSI and MTI are valid, if identity is defined as weak indistinguishability.[8]

The material conditional $\phi_x(b) \rightarrow \phi_x(a)$ fails to be true on Priest's semantics iff $\phi_x(b)$ is true and not also false, and $\phi_x(a)$ is false and not also true. Call any such ϕ a *differentiating formula* for a and b; and when it's of the form Pt, call it a *differentiating predication* for a and b.

A negation $\neg\phi$ is a differentiating formula for a and b only if ϕ is. A conjunction $\phi \wedge \psi$, disjunction $\phi \vee \psi$, or conditional $\phi \rightarrow \psi$ is a differentiating formula for a and b only if at least one of ϕ and ψ is. And a generalisation $\forall x\phi$ or $\forall X\phi$ is a differentiating formula for a and b only if at least one of its instantiations $\phi_x(t), \phi_X(P)$ is. All formulae are recursively built from predications Pt using the connectives and quantifiers just mentioned. So there are no differentiating formulae for a and b unless there are differentiating predications for a and b.

If $\forall X(Xa \leftrightarrow Xb)$ is true, there are no differentiating predications for a and b. Intuitively, any such predication requires a property true and not also false of a, as well as false and not also true of b (or conversely), whereas $\forall X(Xa \leftrightarrow Xb)$ says

[7] *Modus ponens* is: $\phi, \phi \rightarrow \psi \vdash \psi$.

[8] See (Priest, 2014, 30–31) for a more detailed argument. A simpler argument is available given unrestricted λ-conversion or second-order comprehension. Yet these are valid only if every formula determines a corresponding property, which Priest (2014, 24–26) explicitly denies. Comprehension is discussed in §2.6.

that every property is true of both a and b or false of both a and b (and doesn't preclude some properties *also* being true of one and yet false of the other). So if $\forall X(Xa \leftrightarrow Xb)$, there are no differentiating predications and hence no differentiating formulae for a and b. This validates MSI and hence also MTI if the identity sign is defined as weak indistinguishability. One can also run this argument in classical second-order logic, where *modus ponens* is valid; so this also explains why defining identity as weak indistinguishability validates SI and TI in classical second-order logic. Although this argument was couched in relatively informal terms, it's straightforward to check that the model theory described in §2.2 verifies the preceding claims.

2 Strong indistinguishability

Because weak indistinguishability is defined from the material conditional, its properties depend on those of the material conditional. If II concerns weak indistinguishability, this makes the logic of identity, including the validity of principles like SI, dependent on the logic of the material conditional, and in particular on the validity of *modus ponens*.

This section explores two alternative notions of indistinguishability defined without employing a conditional of any kind, material or otherwise. This makes their logical properties independent of the material conditional's. These notions of indistinguishability are conjunctions of purely quantificational notions, whereas weak indistinguishability is a conjunction of partly quantificational and partly conditional notions. (In each case, the conjuncts correspond to the two directions of the biconditional.)

If the first notion of *strong indistinguishability* is used to define the identity sign, TI but not SI is valid. If the second notion of *super-strong indistinguishability* is used to define the identity sign, TI and SI are both valid. So if one these notions is employed instead of weak indistinguishability in II, then TI and perhaps also SI is valid. Both strong and super-strong indistinguishability are classically equivalent to weak indistinguishability, but strictly stronger than it in the present dialetheic setting. In the conclusion (§3), I use this last fact to argue that both strong and super-strong indistinguishability are better candidates to occupy the theoretical role of numerical identity than is weak indistinguishability.

2.1 Definition

To define strong indistinguishability, we need to supplement the language of second-order logic with an irreducibly binary universal second-order quantifier. To understand what this means, consider English generalisations of the form 'all *F*s are *G*s'. In introductory logic, we teach that this should be regimented as:

$$\forall x(Fx \rightarrow Gx)$$

The universal first-order quantifier $\forall x$ here is unary in the following sense: it combines with a single formula—in this case $Fx \rightarrow Gx$—to form a sentence. According to the theory of generalised quantifiers, not all quantifiers are unary.[9] There are also binary quantifiers that combine with two formulae to form sentences. The classic example is 'most': 'most *F*s are *G*s' cannot be adequately regimented using only *F*, *G*, unary quantifiers, and the usual logical connectives. Yet this sentence can be adequately regimented using a binary quantifier that takes both 'Fx' and 'Gx' as arguments. Sentences formed using this quantifier have the form:

$$\text{Most } x(\phi : \psi)$$

where ϕ occupies the quantifier's first argument and ψ its second.

'Most' is not the only binary quantifier. There is also a binary universal quantifier: $\forall x(\phi : \psi)$. We can use this quantifier to regiment 'all *F*s are *G*s' thus:

$$\forall x(Fx : Gx)$$

Unlike the familiar, orthodox regimentation, this one does not employ a seemingly extraneous logical connective $\rightarrow$ that does not appear in the surface form of English. It is thus a purely quantificational regimentation of 'all *F*s are *G*s', whereas the traditional one is partly quantificational and partly conditional.

Indistinguishability is complete sharing of properties: every property of a is a property of b, and conversely. We now have a choice in regimenting this. We can adopt the standard unary regimentation, obtaining weak indistinguishability. Or we can adopt a binary regimentation. Because we are using second-quantifiers to regiment property-talk, this second approach requires a binary second-order universal quantifier. With this to hand, we define strong indistinguishability thus:

$$a \text{ is strongly indistinguishable from } b =_{df} \forall X(Xa : Xb) \wedge \forall X(Xb : Xa)$$

Is there really any difference these two notions of indistinguishability? Does the choice of unary or binary universal quantification have semantic consequences? The rest of this section provides an intuitive reason to think so, before we turn to the formal details in §2.2; we focus on the first-order case for simplicity until then.

Intuitively, the truth of the unary generalisation $\forall x\phi$ requires something of every object in the domain, namely, that it satisfies ϕ. Thus the unary generalisation $\forall x(Fx \rightarrow Gx)$ requires of every object in the domain that it satisfy the logically complex formula $Fx \rightarrow Gx$. By contrast, the truth of the binary generalisation $\forall x(\phi : \psi)$ requires something of only those objects in the domain that satisfy ϕ; it requires nothing of objects that don't satisfy ϕ. Thus the binary generalisation $\forall x(Fx : Gx)$ requires of every satisfier of Fx in the domain that it also satisfy Gx; it requires nothing of the non-satisfiers of Fx, whereas its unary counterpart requires that they also satisfy $Fx \rightarrow Gx$.

In classical semantics, the unary and binary regimentations of 'all *F*s are *G*s' are mutually entailing. Suppose $\forall x(Fx : Gx)$ is true, and hence that every satisfier of Fx

[9] For a comprehensive survey, see (Peters and Westerståhl, 2008). For a shorter overview, see (Westerståhl, 2016) or (Uzquiano, 2018, §3.1).

also satisfies *Gx*. Then every satisfier of *Fx* also satisfies $Fx \to Gx$. And since every non-satisfier of *Fx* satisfies $\neg Fx$, every non-satisfier of *Fx* also satisfies $Fx \to Gx$. So everything satisfies $Fx \to Gx$ and $\forall x(Fx \to Gx)$ is true. Conversely, suppose $\forall x(Fx \to Gx)$ is true, and hence that everything satisfies $Fx \to Gx$. There are two ways of satisfying that formula: by satisfying both *Fx* and *Gx*, or by satisfying $\neg Fx$. So if every object satisfies $Fx \to Gx$, every satisfier of *Fx* also satisfies *Gx*, thereby making $\forall x(Fx : Gx)$ true.

In the dialetheic semantics described in §2.2, this equivalence breaks down. One way of satisfying $Fx \to Gx$ is to satisfy $\neg Fx$. But satisfaction of $\neg Fx$ doesn't preclude satisfaction of *Fx*. So even if some satisfiers of $\neg Fx$ also satisfy *Fx* and yet fail to satisfy *Gx*, they will still satisfy $Fx \to Gx$. The existence of any such objects prevents the binary $\forall x(Fx : Gx)$ from being true, but not the unary $\forall x(Fx \to Gx)$ (though they also make it false, hence, if true, both true and false). So the latter does not entail the former, and the binary is strictly stronger than the unary.

2.2 Semantics

We can make these ideas precise in a model-theoretic setting. I begin by summarising the semantics in chapter two of *One*, and then add an irreducibly binary second-order universal quantifier.

A model is a triple $\langle D_1, D_2, v\rangle$. D_1 is a non-empty set: the domain of first-order quantification. D_2 is any non-empty set of pairs $D = \langle D^+, D^-\rangle$ such that (i) $D^+ \subseteq D_1$, (ii) $D^- \subseteq D_1$, and (iii) $D^+ \cup D^- = D_1$. These pairs represent properties of the individuals (represented by the individuals) in D_1, with D^+ the property's extension (comprising the things that possess it) and D^- its anti-extension (comprising the things that do not possess it). Note that extension and anti-extension may overlap, though they jointly exhaust D_1. The elements of D_2 comprise the domain of second-order quantification. v is a valuation function taking constants c of the object-language to their semantic values $v(c)$ in the model. For singular term constants t, $v(t) \in D_1$. For predicate constants P, $v(P) \in D_2$, with $v^+(P)$ and $v^-(P)$ the extension and anti-extension respectively of $v(P)$; that is, $v(P) = \langle v^+(P), v^-(P)\rangle$. We work with monadic predicates only for simplicity.

Each formula of the language of monadic second-order logic without identity is assigned both truth-conditions ($\Vdash^+$) and falsity-conditions ($\Vdash^-$) in each model $m = \langle D_1, D_2, v\rangle$:

(Pt^+) $m \Vdash^+ Pt$ iff $v(t) \in v^+(P)$

(Pt^-) $m \Vdash^- Pt$ iff $v(t) \in v^-(P)$

($\neg^+$) $m \Vdash^+ \neg\phi$ iff $m \Vdash^- \phi$

($\neg^-$) $m \Vdash^- \neg\phi$ iff $m \Vdash^+ \phi$

($\wedge^+$) $m \Vdash^+ \phi \wedge \psi$ iff $m \Vdash^+ \phi$ and $m \Vdash^+ \psi$

($\wedge^-$) $m \Vdash^- \phi \wedge \psi$ iff $m \Vdash^- \phi$ or $m \Vdash^- \psi$

$(\rightarrow^+)$ $m \Vdash^+ \phi \rightarrow \psi$ iff $m \Vdash^- \phi$ or $m \Vdash^+ \psi$

$(\rightarrow^-)$ $m \Vdash^- \phi \rightarrow \psi$ iff $m \Vdash^+ \phi$ and $m \Vdash^- \psi$

I omit $\vee$ and $\leftrightarrow$ because they're definable from the connectives above and are not required below. For the unary quantifiers, we extend the object-language with a new constant term t_d for each $d \in D_1$ and a new constant predicate P_D for each $D \in D_2$; we also extend the valuation v so that $v(t_d) = d$ and $v(P_D) = D$.[10] Truth- and falsity-conditions are then assigned to quantified sentences thus:

$(\forall^+_{1u})$ $m \Vdash^+ \forall x\phi$ iff, for all $d \in D_1$, $m \Vdash^+ \phi_x(t_d)$

$(\forall^-_{1u})$ $m \Vdash^- \forall x\phi$ iff, for some $d \in D_1$, $m \Vdash^- \phi_x(t_d)$

$(\forall^+_{2u})$ $m \Vdash^+ \forall X\phi$ iff, for all $D \in D_2$, $m \Vdash^+ \phi_x(P_D)$

$(\forall^-_{2u})$ $m \Vdash^- \forall X\phi$ iff, for some $D \in D_2$, $m \Vdash^- \phi_x(P_D)$

I omit the existential quantifier $\exists$ because it's definable from $\forall$ and is not required below. Finally, we define consequence as truth-preservation in all models:

> $\Gamma \models \phi$ iff, for all models m such that $m \Vdash^+ \gamma$ for all $\gamma \in \Gamma$, $m \Vdash^+ \phi$ (where Γ is any collection of closed sentences and ϕ is a closed sentence).

If the identity sign is defined as weak indistinguishability, this semantics validates MSI and MTI but not SI and TI. That is:

$$\begin{array}{ll} a = b \models \phi_x(b) \rightarrow \phi_x(a) & \quad a = b, \phi_x(b) \not\models \phi_x(a) \\ a = b \models b = c \rightarrow a = c & \quad a = b, b = c \not\models a = c \end{array}$$

We'll see models that verify the right-hand results in §2.4 and §2.5.

We now add binary universal quantifiers, thus:

$(\forall^+_{1b})$ $m \Vdash^+ \forall x(\phi : \psi)$ iff, for all $d \in D_1$ such that $m \Vdash^+ \phi_x(t_d)$, $m \Vdash^+ \psi_x(t_d)$

$(\forall^-_{1b})$ $m \Vdash^- \forall x(\phi : \psi)$ iff, for some $d \in D_1$ such that $m \Vdash^+ \phi_x(t_d)$, $m \Vdash^- \psi_x(t_d)$

$(\forall^+_{2b})$ $m \Vdash^+ \forall X(\phi : \psi)$ iff, for all $D \in D_2$ such that $m \Vdash^+ \phi_X(P_D)$, $m \Vdash^+ \psi_x(P_D)$

$(\forall^-_{2b})$ $m \Vdash^- \forall X(\phi : \psi)$ iff, for some $D \in D_2$ such that $m \Vdash^+ \phi_X(P_D)$, $m \Vdash^- \psi_x(P_D)$

[10] Rather than extending the language with new constants, we could relativise $\Vdash^+$ and $\Vdash^-$ to assignments of elements of D_1 and D_2 to the first- and second-order variables respectively. The clauses for the quantifiers would then concern truth- and falsity under all assignments differing at most over the variable bound by the quantifier, rather than all sentences in the extended language. For present purposes, nothing of philosophical or technical significance turns on this choice. The presentation in the body text is chosen only because it matches Priest's own.

It is instructive to compare these truth-conditions for binary generalisations with those of corresponding unary generalisations. I focus on the first order quantifiers for simplicity, though what I say holds for the second-order quantifiers *mutatis mutandis* too.

Consider unary generalisations of the form $\forall x(\phi \rightarrow \psi)$ and their binary counterparts $\forall x(\phi : \psi)$. The latter has truth-condition $(\forall^{+}_{1b})$. The former has the following truth-condition, by $(\forall^{+}_{1u})$ and $(\rightarrow^{+})$:

$$m \Vdash^{+} \forall x(\phi \rightarrow \psi) \text{ iff, for all } d \in D_1,\ m \Vdash^{-} \phi_x(t_d) \text{ or } m \Vdash^{+} \psi_x(t_d)$$

Say that *d satisfies* ϕ in m iff $m \Vdash^{+} \phi_x(t_d)$, and *anti-satisfies* ϕ in m iff $m \Vdash^{-} \phi_x(t_d)$. In this terminology, the truth of $\forall x(\phi \rightarrow \psi)$ in m requires that every single $d \in D_1$ either anti-satisfies ϕ in m or satisfies ψ in m (or both). By contrast, the truth of $\forall x(\phi : \psi)$ in m requires only that every satisfier of ϕ in m is also a satisfier of ψ in m. As a result, binary generalisations entail their unary counterparts but not conversely:

$$\forall x(\phi : \psi) \models \forall x(\phi \rightarrow \psi) \qquad \forall x(\phi \rightarrow \psi) \not\models \forall x(\phi : \psi)$$

To see why the second entailment fails, consider the following countermodel:

$D_1 = \{0\}$

$D_2 = \{P, Q\}$

$P^{+} = \{0\} \quad P^{-} = \{0\}$
$Q^{+} = \emptyset \quad Q^{-} = \{0\}$

$v(F) = P$

$v(G) = Q$

In this model $0 \in v^{-}(F)$ and hence 0 anti-satisfies *Fx*. Since 0 is the only element of D_1, $\forall x(Fx \rightarrow Gx)$ is true. Since $0 \in v^{+}(F)$, 0 also satisfies *Fx*. But since $0 \notin \emptyset$, it follows that $0 \notin v^{+}(G)$, 0 doesn't satisfy *Gx*, and so $\forall x(Fx : Gx)$ is not true. So this is a countermodel to $\forall x(\phi \rightarrow \psi) \models \forall x(\phi : \psi)$. Since $\forall x(\phi : \psi) \models \forall x(\phi \rightarrow \psi)$, however, the binary generalisation is strictly stronger than its unary counterpart.

This difference in logical strength vanishes in classical semantics. We can obtain classical semantics from the present semantics by requiring that D^{+} and D^{-} are disjoint, for all $D \in D_2$. This precludes the countermodel above. Moreover, failure to satisfy ϕ coincides with anti-satisfaction of ϕ in this version of the semantics. We can then show that $\forall x(\phi \rightarrow \psi) \models \forall x(\phi : \psi)$ in classical semantics. Suppose $m \Vdash^{+} \forall x(\phi \rightarrow \psi)$ and hence that, for all $d \in D_1$, d anti-satisfies ϕ or satisfies ψ. Consider an arbitrary $d \in D_1$ that satisfies ϕ. Since anti-satisfaction coincides with failure to satisfy, d does not anti-satisfy ϕ. So d must satisfy ψ. Since d was arbitrary, every $d \in D_1$ that satisfies ϕ also satisfies ψ, and hence $m \Vdash^{+} \forall x(\phi : \psi)$. Since m was arbitrary $\forall x(\phi \rightarrow \psi) \models \forall x(\phi : \psi)$.

The lesson is that the choice between unary and binary regimentations of 'all *F*s are *G*s' doesn't matter in classical semantics. In dialetheic semantics—or any other

many-valued semantics—differences can emerge. In the present setting, the binary regimentation entails the unary regimentation but not conversely. Although we've focussed on first-order quantifiers thus far, analogous results hold for the second-order quantifiers, to which we now turn.

2.3 Logical strength

The choice between unary and binary regimentations of 'all *F*s are *G*s' matters when regimenting complete sharing of properties: every property of a is a property of b, and conversely. The unary regimentation is weak indistinguishability whereas the binary regimentation is strong indistinguishability.

Second-order analogs of the facts discussed in the preceding section hold for essentially the same reasons as in the first-order case. In particular:

$$\forall X(\phi : \psi) \models \forall X(\phi \rightarrow \psi) \qquad \forall X(\phi \rightarrow \psi) \not\models \forall X(\phi : \psi)$$

Here's a countermodel to show that the unary regimenation does not entail its binary counterpart:

$D_1 = \{0, 1\}$

$D_2 = \{Q\}$

$Q^+ = \{0\} \quad Q^- = \{0, 1\}$

$v(a) = 0$

$v(b) = 1$

$v(F) = Q$

In this model, $v(a)$ and $v(b)$ both belong to $v^-(F)$ because both belong to Q^-. So each of $Fa \rightarrow Fb$ and $Fb \rightarrow Fa$ is true (because each antecedent is false). Since Q is the only element of D_2 and $v(F) = Q$, the same holds for every predicate P that can be added to the language with $v(P) \in D_2$. So each of $\forall X(Xa \rightarrow Xb)$ and $\forall X(Xb \rightarrow Xa)$ is true. So it's true in this model that a is weakly indistinguishable from b. Yet since $v(a)$ but not $v(b)$ belongs to $v^+(F)$, i.e. Q^+, $\forall X(Xa : Xb)$ is not true. So it's not true in this model that a is strongly indistinguishable from b, and so $\forall X(\phi \rightarrow \psi) \not\models \forall X(\phi : \psi)$. Because binary second-order generalisations entail their unary counterparts but not conversely, strong indistinguishability entails weak indistinguishability but not conversely.

One striking feature of this model is that its second-order domain has only one member. Many formulae do not define properties in this model. So it does not verify all instances of the second-order comprehension schema:

$$\exists Y \forall z(Yz \leftrightarrow \phi) \text{ (where } Y \text{ is not free in } \phi)$$

One might wonder whether countermodels to $\forall X(\phi \to \psi) \models \forall X(\phi : \psi)$ are available only because of this unusually permissive conception of the second-order domain. Perhaps the logical difference between weak and strong indistinguishability vanishes if we require that the second-order domain is much fuller. In fact, this is not the case; but since it's a somewhat involved matter, I postpone discussion to §2.6 to avoid getting sidetracked now.

To get a better feel for the logical relations between weak and strong indistinguishability, let's look more closely at the logical form of the clauses for the binary quantifiers, and compare them with their unary counterparts. What exactly is the logical form of, e.g., 'for all $d \in D_1$ such that $m \Vdash^+ \phi_x(t_d)$, $m \Vdash^+ \psi_x(t_d)$'? There are a couple of options.

The first option employs a primitive binary universal quantifier in the metalanguage. Using Λ for this quantifier to differentiate it from the object-language $\forall$, we can rewrite $(\forall^+_{1b})$ and $(\forall^+_{2b})$ as:

$$m \Vdash^+ \forall x(\phi : \psi) \text{ iff } \Lambda d(d \in D_1 \text{ and } m \Vdash^+ \phi_x(t_d) : m \Vdash^+ \psi_x(t_d))$$

$$m \Vdash^+ \forall X(\phi : \psi) \text{ iff } \Lambda D(D \in D_2 \text{ and } m \Vdash^+ \phi_X(P_D) : m \Vdash^+ \psi_X(P_D))$$

On this approach, the logic of the object-language binary quantifiers turns on that of the meta-language binary quantifier. In particular, §2.4's argument for the validity of TI (if identity is defined as strong indistinguishability) depends on the validity of the following metalinguistic form of binary universal instantiation:

$$\Lambda x(\phi : \psi), \phi_x(a), \text{therefore } \psi_x(a).$$

This principle allows us to instantiate the second condition ψ with arbitrary satisfiers of the first condition ϕ. Although appealing to metalinguistic binary quantifiers is non-standard, I see no problem with it in principle, or with the validity of this particular rule of inference in the metalanguage. The goal of the present model theory is to use classical mathematics to characterise a non-classical consequence relation for the object language. And the rule just presented is surely correct for binary universal quantifiers in a classical setting. But it is worth noting that we do not have to invoke such non-standard resources.

The second option employs a material conditional in the metalanguage. Using $\Rightarrow$ for this conditional, to differentiate it from the object-language $\to$, we can rewrite $(\forall^+_{1b})$ and $(\forall^+_{2b})$ as:

$$m \Vdash^+ \forall x(\phi : \psi) \text{ iff, for all } d\big((d \in D_1 \text{ and } m \Vdash^+ \phi_x(t_d)) \Rightarrow m \Vdash^+ \psi_x(t_d)\big)$$

$$m \Vdash^+ \forall X(\phi : \psi) \text{ iff, for all } D\big((D \in D_2 \text{ and } m \Vdash^+ \phi_X(P_D)) \Rightarrow m \Vdash^+ \psi_X(P_D)\big)$$

We can also rewrite the truth-conditions for the corresponding unary generalisations as:

$$m \Vdash^+ \forall x(\phi \to \psi) \text{ iff, for all } d\big((d \in D_1 \text{ and } m \nVdash^- \phi_x(t_d)) \Rightarrow m \Vdash^+ \psi_x(t_d)\big)$$

$m \Vdash^{+} \forall X(\phi \rightarrow \psi)$ iff, for all $D\big((D \in D_2$ and $m \nVdash^{-} \phi_X(P_D)) \Rightarrow m \Vdash^{+} \psi_X(P_D)\big)$

We can now see the underlying difference between binary generalisations and their unary counterparts as essentially that between the following:

Binary: it satisfies $\phi \Rightarrow$ it satisfies ψ.

Unary: it does not anti-satisfy $\phi \Rightarrow$ it satisfies ψ.

In classical semantics, anti-satisfaction coincides with failure of satisfaction, and so failure of anti-satisfaction coincides with satisfaction. This makes Binary and Unary equivalent. In the present dialetheic semantics, failure to anti-satisfy entails satisfaction; so Binary entails Unary. But satisfaction does not entail failure to anti-satisfy, since formulae can be both true and false; so Unary does not entail Binary.

To close this section, let's apply these ideas to indistinguishability. a and b are strongly indistinguishable when every property with a in its extension also has b in its extension, and conversely. By contrast, a and b are weakly indistinguishable when every property with a not in its anti-extension has b in its extension, and conversely. Because a property without a in its anti-extension must have a in its extension, strong indistinguishability entails weak indistinguishability. But because a property can have a in its extension whilst also having a in its anti-extension, weak indistinguishability does not entail strong indistinguishability. Were satisfaction and anti-satisfaction not just exhaustive but exclusive, however, then weak and strong indistinguishability would be coextensive.

2.4 Transitivity

Although weak indistinguishability is not transitive, strong indistinguishability is transitive. It follows that TI is valid if the identity sign is defined by strong but not weak indistinguishability.

It may be helpful to begin with a countermodel to the transitivity of weak indistinguishability:

$D_1 = \{0, 1, 2\}$

$D_2 = \{Q\}$
$Q^{+} = \{0, 1\} \quad Q^{-} = \{1, 2\}$

$v(a) = 0$

$v(b) = 1$

$v(c) = 2$

In this model, a is weakly indistinguishable from b because every property (i.e. Q) in D_2 with $v(a)$ in its extension also has $v(b)$ in its extension, and conversely. This model also makes b weakly indistinguishable from c because every property with $v(b)$ in its anti-extension also has $v(c)$ in its anti-extension, and conversely.[11] The model does not make a weakly indistinguishable from c, however, because although $v(a) \in Q^+$, we have neither $v(c) \in Q^+$ nor $v(a) \in Q^-$, which makes $P_Q a \rightarrow P_Q c$ false and not also true. So this model shows that weak indistinguishability is not transitive and invalidates TI if the identity sign is defined by weak indistinguishability.

Although a is strongly indistinguishable from b in this model, b is not strongly indistinguishable from c; so the model also shows how weak indistinguishability fails to entail strong indistinguishability. b and c are not strongly indistinguishable because although Q is a property such that $v(b) \in Q^+$, it's not the case that $v(c) \in Q^+$. So this isn't a countermodel to the transitivity of strong indistinguishability; in fact, there are no such countermodels.

Strong indistinguishability is transitive just in case the following holds:

$$\forall X(Xa : Xb) \wedge \forall X(Xb : Xa), \forall X(Xb : Xc) \wedge \forall X(Xc : Xb) \models \forall X(Xa : Xc) \wedge \forall X(Xc : Xa)$$

That holds if both of the following do:

$$\forall X(Xa : Xb), \forall X(Xb : Xc) \models \forall X(Xa : Xc)$$

$$\forall X(Xc : Xb), \forall X(Xb : Xa) \models \forall X(Xc : Xa)$$

Since those have the same form, it suffices to examine the first. We can show that it holds as follows. Suppose $\forall X(Xa : Xb)$ and $\forall X(Xb : Xc)$ are both true in an arbitrary model m. That is:

For all $D \in D_2$ such that $v(a) \in D^+$, $v(b) \in D^+$.

For all $D \in D_2$ such that $v(b) \in D^+$, $v(c) \in D^+$

Consider an arbitrary $D \in D_2$ such that $v(a) \in D^+$. By the first clause, $v(b) \in D^+$. So D is such that $v(b) \in D^+$. So by the second clause, $v(c) \in D^+$. Since D was arbitrary: for all $D \in D_2$ such that $v(a) \in D^+$, $v(c) \in D^+$. But then $m \Vdash^+ \forall X(Xa : Xc)$. And since m was arbitrary: $\forall X(Xa : Xb), \forall X(Xb : Xc) \models \forall X(Xa : Xc)$. It follows that strong indistinguishability is transitive. So if the identity sign is defined as strong indistinguishability, then TI is valid: numerical identity is transitive, dialetheism and the non-transitivity of weak indistinguishability notwithstanding.

[11] Were $v(b)$ but not $v(c)$ in the anti-extension of some property Q^*, the conditional $P_{Q^*} c \rightarrow P_{Q^*} b$ would not be true, and so b would not be weakly indistinguishable from c. Note also that although the model makes a, b and b, c weakly indistinguishable, it also makes them not weakly indistinguishable. That's because the following are both true and false in the model: $P_Q a \rightarrow P_Q b$, $P_Q b \rightarrow P_Q c$. The first is false because $v(a) \in P^+$ and $v(b) \in P^-$. The second is false because $v(b) \in P^+$ and $v(c) \in P^-$.

2.5 Substitutivity

We've seen that if identity is defined as strong indistinguishability, then TI is valid. Interestingly, however, this does not validate the full-strength substitution principle SI, of which TI is an instance. I first explain why, and then introduce a notion of super-strong indistinguishability that validates SI when used to define identity.

To see why SI fails, consider this instance of it:

$$a = b, \neg Fa \vdash \neg Fb$$

If the identity sign is defined as strong (or weak) indistinguishability, the following model invalidates that sequent:

$$D_1 = \{0, 1\}$$

$$D_2 = \{Q\}$$

$$Q^+ = \{0, 1\} \quad Q^- = \{0\}$$

$$v(a) = 0$$

$$v(b) = 1$$

$$v(F) = Q$$

In this model, a is strongly (hence also weakly) indistinguishable from b because the only property Q in D_2 has both $v(a)$ and $v(b)$ in its extension Q^+. The model makes $\neg Fa$ true because $v(a)$ is in the anti-extension Q^- of $v(F)$. But because $v(b)$ is not in Q^-, the model does not make $\neg Fb$ true and therefore invalidates the sequent displayed above.

We can define a stronger notion of indistinguishability that validates SI when used to define the identity sign. Whereas strong indistinguishability requires only that every property of a is a property of b and conversely, this notion requires also that every property not of a is also not a property of b and conversely. Formally:

> a is super-strongly indistinguishable from $b =_{df} \forall X(Xa : Xb) \wedge \forall X(Xb : Xa) \wedge \forall X(\neg Xa : \neg Xb) \wedge \forall X(\neg Xb : \neg Xa)$

Semantically, a model makes a super-strongly indistinguishable from b iff both the following hold:

> For every $D \in D_2$ such that $v(a) \in D^+$, $v(b) \in D^+$, and conversely.
>
> For every $D \in D_2$ such that $v(a) \in D^-$, $v(b) \in D^-$, and conversely.

In short: $v(a)$ and $v(b)$ belong to the extensions and anti-extensions of exactly the same properties.

Super-strong indistinguishability is the conjunction of strong indistinguishability with $\forall X(\neg Xa : \neg Xb) \wedge \forall X(\neg Xb : \neg Xa)$. So super-strong indistinguishability entails strong indistinguishability, and TI is valid when identity is defined

as super-strong indistinguishability. Unlike strong indistinguishability, however, this definition of identity also validates SI.

To get a feel for why this definition of identity validates SI, consider the model described at the beginning of this section. In that model, a is not super-strongly indistinguishable from b because although $v(a)$ and $v(b)$ are both in Q^+, only $v(a) \in Q^-$. Were we to make them super-strongly indistinguishable, by either removing $v(a)$ from Q^- or adding $v(b)$ into Q^-, we'd no longer have a counter-model to the initial sequent; for removing $v(a)$ from Q^- would make $\neg Fa$ untrue, and adding $v(b)$ to Q^- would make $\neg Fb$ true.

A general argument is also available; it's a variant of the argument for MTI and MSI at the end of §1. If identity is defined as super-strong indistinguishability, then a countermodel to SI is any model that satisfies both the following:

> a is super-strongly indistinguishable from b: $v(a)$ and $v(b)$ belong to exactly the same extensions and anti-extensions of all $D \in D_2$.
>
> Some ϕ is a differentiating formula for a and b in the following sense:[12]
>
> > $\phi_x(b)$ is true and $\phi_x(a)$ is false and not true.

Unless ϕ is a predication, a negation $\neg\phi$ is a differentiating formula for a and b only if ϕ is; we saw above that negations of predications may be differentiating formulae even when the predication itself isn't. A conjunction $\phi \wedge \psi$, disjunction $\phi \vee \psi$, or material conditional $\phi \rightarrow \psi$ is a differentiating formula for a and b only if at least one of ϕ and ψ is. And a generalisation $\forall x\phi$ or $\forall X\phi$ is a differentiating formula for a and b only if at least one of its instantiations $\phi_x(t), \phi_X(P)$ is. All formulae are recursively built from predications Pt using the connectives and quantifiers just mentioned. So there are no differentiating formulae for a and b unless some predication or its negation is a differentiating formula for a and b. But if a is super-strongly indistinguishable from b, no predication Pt or its negation $\neg Pt$ is a differentiating formula for a and b, regardless of which $D \in D_2$ is $v(P)$. So if a is super-strongly indistinguishable from b, there are no differentiating formulae for a and b whatsoever. SI is therefore valid if identity is defined as super-strong indistinguishability.

I now argue that super-strong indistinguishability provides a better dialetheic regimentation of the intuitive notion of complete sharing of properties than does strong indistinguishability. §3 offers a related argument for the conclusion that super-strong indistinguishability is a better candidate to occupy the theoretical role of numerical identity than is strong indistinguishability.

Indistinguishable objects are not distinguished by their properties. In classical metaphysics, that's ruled out by possession of the same properties. In dialetheic

[12] This is a weaker notion of differentiating formula than that required in §1, since our goal is to invalidate SI rather than MSI. Not all differentiating formulae in the present sense are differentiating formulae in the earlier sense.

metaphysics, however, objects may possess the same properties and yet be distinguished by them. That happens whenever one but not the other object both has and lacks the property. An adequate regimentation of indistinguishability should ensure that indistinguishable objects are not distinguished by which properties they possess, or by which properties they lack. Neither weak nor strong indistinguishability ensures this, whereas super-strong indistinguishability does. Super-strong indistinguishability therefore provides a better dialetheic regimentation of indistinguishability than does either weak or strong indistinguishability. If numerical identity is indistinguishability, as II says, then numerical identity is super-strong indistinguishability, and so SI and TI are valid.

2.6 Comprehension

As noted in §2.4, the models discussed so far have very few elements in their second-order domains. Many formulae do not define properties in such models, and so instances of of the following second-order comprehension schema are not valid:

$$\exists Y \forall z (Yz \leftrightarrow \phi) \text{ (where } Y \text{ is not free in } \phi)$$

This section shows that the difference in logical strength between weak and strong indistinguishability is not an artefact of this unusually permissive conception of the second-order domain. We will see, however, that this is relevant to the relative logical strength of strong and super-strong indistinguishability. In particular, a natural constraint on models that serves to validate comprehension makes strong and super-strong indistinguishability mutually entailing.

Recall the following model from §2.4:

$$D_1 = \{0, 1, 2\}$$

$$D_2 = \{Q\}$$
$$Q^+ = \{0, 1\} \quad Q^- = \{1, 2\}$$

$$v(a) = 0$$

$$v(b) = 1$$

$$v(c) = 2$$

Because the pairs a, b and b, c but not a, c are weakly indistinguishable in this model, it invalidates TI and SI when identity is defined as weak indistinguishability. Because b, c are weakly but not strongly or super-strongly indistinguishable, the model shows that weak does not entail strong or super-strong. And because a, b are strongly but not super-strongly indistinguishable, it shows that strong does entail super-strong.

The goal now is to modify this model so as to verify each instance of comprehension without disrupting the pattern of weak indistinguishability, and whilst

retaining pairs that are weakly but not strongly or super-strongly indistinguishable. This will show that the following are compatible with the validity of second-order comprehension: (i) defining identity as weak indistinguishability invalidates SI and TI; (ii) strong and super-strong indistinguishability are strictly stronger than weak indistinguishability. In the kind of model we end up with, however, strong and super-strong indistinguishability are mutually entailing; that particular difference in logical strength is dependent on Priest's permissive conception of the second-order domain. The technique employed is adapted from (Priest, 2010a, §3.2).

First, some terminology. For each formula ϕ with at most one free variable x, and each model m, set:

$v_m^+(\phi) = \{d \in D_1 : m \Vdash^+ \phi_x(t_d)\}$ (i.e. the m-extension of ϕ)

$v_m^-(\phi) = \{d \in D_1 : m \Vdash^- \phi_x(t_d)\}$ (i.e. the m-anti-extension of ϕ)

$v_m(\phi) = \langle v_m^+(\phi), v_m^-(\phi)\rangle$

We will proceed by successively adding to D_2. Rename the second-order domain in the model above D_2^0 and let model m^i be $\langle D_1, D_2^i, v\rangle$; so the model above is m^0. Let the language of m^i be the result of extending the object-language with a new predicate P_D for each $D \in D_2^i$, and extend the valuation v so that $v(P_D) = D$. The process of addition is then defined by:

> $D_2^i = D_2^{i-1} \cup \{v_{m^{i-1}}(\phi) : \phi$ is a formula of the language of m^{i-1} with at most one free variable$\}$.

Because D_1 is finite, the process terminates after finitely many steps in that $D_2^i = D_2^{i+1}$. Were D_1 infinite, we would need an additional clause for limit ordinals i to ensure that the process terminates:

$$D_2^i = \bigcup_{j<i} D_2^j.$$

Let M be the model at which the process terminates.

Pairs that are not weakly/strongly/super-strongly indistinguishable in m^0 remain so in all later m^i including M; for later models only add to D_2, and so leave all counterexamples to weak/strong/super-strong indistinguishability in place. So we only need to check that weakly and super-strongly indistinguishable pairs in m^0 also remain so in all later m^i. I discuss the behaviour of pairs that are strongly but not super-strongly indistinguishable shortly.

In this particular model, the super-strongly indistinguishable pairs are just a, a, and b, b, and c, c. Clearly, no way of adding to D_2 can provide counterexamples to those. So they're all super-strongly indistinguishable in M.

Let's turn to weakly indistinguishable pairs. An arbitrary pair α, β are not weakly indistinguishable in m^i iff, for some P_D in the language of m^i, the predication $P_D x$ strongly differentiates α from β in the following sense:

$P_D\alpha$ is true and not also false in m^i

$P_D\beta$ is false and not also true in m^i

or conversely. Each element of D_2^{i+1} has the same extension and anti-extension as some formula in the language of m^i (or some other earlier m^j). So each predicate in the language of m^{i+1} has the same extension and anti-extension as some formula in the language of m^i (or some other earlier m^j). So there is a strongly differentiating predication for α and β in the language of m^{i+1} iff there is a strongly differentiating formula for α and β in the language of m^i (or some other earlier m^j). We saw at the end of §1 that there are no strongly differentiating formulae for α and β in the language of any m if there are no strongly differentiating predications α and β in the language of m. So if there are no strongly differentiating predications for α and β in the language of m^0, there are no strongly differentiating predications for α and β in the language of any later m^i, including M. Since α and β were arbitrary, the pattern of weak indistinguishability in m^0 is therefore also present in M.

It remains only to see that each instance of comprehension is true in M. Since $v_M(\phi) \in D_2^M$, for all ϕ in the language of M, each instance of comprehension is true in M. To see this, instantiate the second-order quantifier for a predicate $P_{v_M(\phi)}$ in the language of M such that $v(P_{v_M(\phi)}) = v_M(\phi)$:

$$\forall z(P_{v_M(\phi)}z \leftrightarrow \phi)$$

Because ϕ and $P_{v_M(\phi)}$ have the same extension and anti-extension in M, each instance of this schema is true in M. That suffices for the second-order existential generalisation of each instance of the schema to be true in M too.

Putting these pieces together, some countermodels to the transitivity of weak indistinguishability also verify each instance of comprehension. These models also contain weakly indistinguishable pairs that are not strongly or super-strongly indistinguishable, which shows that this difference in logical strength is independent of the validity or otherwise of comprehension.

Say that a model m is *comprehensive* iff, for all ϕ in the language of m, $v_m(\phi) \in D_2$. Each instance of comprehension is true in each comprehensive model. So restricting the models to the comprehensive models validates each instance of comprehension.[13] It also makes strong and super-strong indistinguishability mutually entailing. To see this, consider any strongly indistinguishable pair a, b. Because they're strongly indistinguishable, they're in the extensions of exactly the same properties. We can show that if the model is comprehensive, a and b lack exactly the same properties too, and hence that they're super-strongly indistinguishable. So suppose for *reductio* that they don't lack exactly the same properties, that is, for some $D \in D_2$, $v(a) \in D^-$ and $v(b) \notin D^-$. Then $m \Vdash^+ \neg P_D b$, and so $b \in v_m^+(\neg P_D x)$. Since m is comprehensive $v_m(\neg P_D x) \in D_2$. But then, because a and b have exactly the same properties, $v(a) \in v_m^+(\neg P_D x)$ too. So $m \Vdash^+ \neg P_D a$, which means

[13] We can also validate λ-conversion over the class of comprehensive models, by setting $v(\lambda x.\phi) = v_m(\phi)$. Although other ways of validating comprehension and λ-conversion are available, only the restriction to comprehensive models ensures that λ commutes with negation.

$m \Vdash^{-} P_D a$ and hence $v(a) \notin D^{-}$, contrary to the initial supposition. By *reductio*, a and b lack exactly the same properties in comprehensive models where they possess exactly the same properties. That is: strong indistinguishability entails super-strong indistinguishability in comprehensive models.

The restriction to comprehensive models is a natural way to validate comprehension. It also makes strong and super-strong indistinguishability mutually entailing. Since defining identity as super-strong indistinguishability validates SI, defining identity as strong indistinguishability validates SI over the comprehensive models. The apparent weakness of strong indistinguishability therefore depends on a permissive conception of the domain of second-order quantification; I take no stand here on whether that conception is correct or not.

3 Conclusion

We've seen three classically equivalent notions of indistinguishability with different logical properties in the present dialetheic setting. Super-strong entails strong, which entails weak, but none of the converse entailments hold (except for that from strong to super-strong within the comprehensive models). II says that identity is indistinguishability. This allows us to introduce the identity sign with its intended interpretation into an identity-free second-order language by defining it as indistinguishability. Different choices about the regimentation of indistinguishability then yield different logical properties for identity. If indistinguishability is weak indistinguishability, neither SI nor TI is valid even in paradigmatically extensional contexts (§1). If indistinguishability is strong indistinguishability, TI but not SI is valid (§§2.4–2.5). And if indistinguishability is super-strong indistinguishability, both TI and SI are valid (§2.5).

Which of these views is correct, given II? Is identity weak indistinguishability, or strong, or super-strong? Is this dispute even substantive? Suppose we all agreed to stop using the words 'numerical identity', and to replace them with words for the various different relations of indistinguishability instead. Would anything be lost? Would any facts be left out? I think so.

There is a deep connection between numerical identity and objecthood. Reality comprises, at least in part, an array of objects instantiating properties and standing in relations. This supply of objects determines how fine-grained reality's distinctions can be. Given a collection of many objects, there can be distinctions amongst them. For example, the red ones on one side and the rest on the other. If many red objects remain, further distinctions can be made within them. And so on. But when the collection comprises just one single object, no more fine-grained distinctions can be made. When a is identical to b, no distinction treats them differently, whatever the supply of distinctions may be. Identical objects fall on exactly the same side(s) of each distinction; that's what it is for no more fine-grained distinction to be possible between them. Numerical identity thus limits how fine-grained reality can be.

Let me be clear about exactly what this connection between identity and fineness of grain requires, to emphasise its compatibility with dialetheism. It does not require—at least, not without additional assumptions incompatible with dialetheism—that identical objects never lie on opposite sides of the same distinction. Dialetheists should reject that idea because whenever a is both F and not F, a lies on opposite sides of the F/non-F distinction from itself. Rather, what's required is that whenever a is identical to b, any distinction with a on a given side also has b on that side. That's what it is for no distinction to differentiate between a from b, or to cut more finely than them. Dialetheism is entirely compatible with that.

This connection between objecthood, identity, and fineness of grain generates a core theoretical role for numerical identity: it is the most demanding and restrictive form of indistinguishability. A relation that permits differences of any sort between its relata is not numerical identity because it does not prevent there from being more fine-grained distinctions between its relata. Of the relations we've examined, only super-strong indistinguishability is even a candidate to occupy this role. Weak and strong both permit violations of SI because they permit distinctions between their relata over which properties they possess and lack. When a is only weakly indistinguishable from b, a may have or lack properties that b does not. When a is only strongly distinguishable from b, a may lack properties that b does not. In both cases, the pattern of possession and lack differentiates a from b, showing that they are not numerically identical in the sense just outlined. Super-strong indistinguishability permits no such differences between its relata, and as a result validates both SI and TI. The proper treatment of identity in dialetheic metaphysics therefore identifies it with super-strong indistinguishability.

References

Cobreros, P., Egré, P., Ripley, D., and van Rooij, R. (2013). Identity, leibniz's law, and non-transitive reasoning. *Metaphysica*, 14(2):253–264.

Cobreros, P., Egré, P., Ripley, D., and van Rooij, R. (2014). Priest's motorbike and tolerant identity. In Ciuni, R., Wansing, H., and Willkommen, C., editors, *Recent Trends in Philosophical Logic*, pages 75–84. Springer.

Forrest, P. (2016). The identity of indiscernibles. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, winter 2016 edition. `https://plato.stanford.edu/archives/win2016/entries/identity-indiscernible/`.

Hawley, K. (2009). Identity and indistiguishability. *Mind*, 118(469):101–119.

Hawthorne, J. (2003). Identity. In Loux, M. J. and Zimmerman, D. W., editors, *The Oxford Handbook of Metaphysics*, chapter 4. OUP. Reprinted as (Hawthorne, 2006, ch1).

Hawthorne, J. (2006). *Metaphysical Essays*. OUP.

Jones, N. K. (2018). How to unify. *Ergo: An Open Access Journal of Philosophy*. Forthcoming.

Noonan, H. and Curtis, B. (2018). Identity. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, summer 2018 edition. `https://plato.stanford.edu/archives/sum2018/entries/identity/`.

Peters, S. and Westerståhl, D. (2008). *Quantifiers in Language and Logic*. OUP.

Priest, G. (2009). A case of mistaken identity. In Lear, J. and Oliver, A., editors, *The Force of Argument*. Routledge.

Priest, G. (2010a). Contradiction and the structure of unity. In Jiang, Y., editor, *Analytic Philosophy in China 2009*, pages 35–42. Zheijiang University Press. Accessed via `http://grahampriest.net/?ddownload=1193`.

Priest, G. (2010b). Non-transitive identity. In *Cuts and Clouds: Vagueness, Its Nature and Its Logic*, chapter 23. OUP.

Priest, G. (2014). *One: Being an Investigation into the Unity of Reality and of Its Parts, Including the Singular Object which is Nothingness*. OUP.

Uzquiano, G. (2018). Quantifiers and quantification. In *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, fall 2018 edition. `https://plato.stanford.edu/archives/fall2018/entries/quantification/`.

Westerståhl, D. (2016). Generalized quantifiers. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, winter 2016 edition. `https://plato.stanford.edu/archives/win2016/entries/generalized-quantifiers/`.