Philosophy and
Phenomenological Research

Philosophy and Phenomenological Research
doi: 10.1111/phpr.12587

Classical Opacity*

MICHAEL CAIE
University of Toronto

JEREMY GOODMAN
University of Southern California

HARVEY LEDERMAN
Princeton University

1 Introduction

It is an apparent truism that, for any things x and y, if x and y are *identical*, then x and y have the same properties. For if x and y are identical they are one and the same thing, and so it seems whatever properties the one has the other must have as well. In this paper we will explore views that deny this apparent truism. We begin with familiar examples like the following:

Hesperus/Phosphorus: Hesperus is Phosphorus. But while the ancients knew that Hesperus was visible at night, the ancients did not know that Phosphorus was visible at night.

This example presents a putative false instance of the following schematic principle:

Substitution $a = b \rightarrow (\varphi \leftrightarrow \varphi[b/a])$.[1]

We will call false instances of Substitution cases of *opacity*. Given the existence of cases of opacity, it follows, given classical quantificational logic, that there are false instances of:

Quantified Substitution $\forall x \forall y(x = y \rightarrow (\varphi \leftrightarrow \varphi[y/x]))$.

* The authors thank Andrew Bacon, Cian Dorr, Peter Fritz, Dmitri Gallow, Agustín Rayo, Erica Schumener, James Shaw for discussion and comments on drafts of this material, as well as audiences at NYU, Carnegie Mellon, UT Austin, Toronto, Birmingham, and Princeton. The authors contributed equally to the ideas of all parts of the paper, but with the exception of Appendix C.2.2, MC and HL wrote the appendices.

[1] Instances of this schema are obtained by replacing φ with a declarative sentence s, replacing a and b with proper names m and n, and replacing $\varphi[b/a]$ with a sentence obtained from s by replacing one or more occurrences of m that are not within quotation marks with an occurrence of n.

And, given classical quantification logic and other minimal logical assumptions (which we will spell out shortly), it also follows that there are false instances of:

Leibniz's Law $a = b \rightarrow \forall X\,(Xa \leftrightarrow Xb)$.

In this paper we will explore the idea that even if Substitution has false instances, classical quantification theory (and the aforementioned minimal background logical assumptions) should be upheld. If this is right, and Substitution does have false instances, then the apparent truism with which we began is false. There are individuals x (Hesperus) and y (Phosphorus) that are identical despite there being a property X (being known by the ancients to be visible at night) that x has and y does not have.

There is of course a vast literature devoted to reconciling Substitution with putative cases of opacity.[2] Neo-Russellians, for example, argue that despite its *prima facie* plausibility the second sentence of Hesperus/Phosphorus is false.[3] And some contextualists have argued that while uses of this sentence may be true, these uses involve equivocation and the sentence is false provided there is no mid-sentence shift in context.[4]

But others (for example, Fregeans) are committed to holding that there are cases of opacity.[5] In this paper we will assume that they are correct. We make this assumption in an exploratory spirit. Our goal here is to study what the logic of identity should be if Substitution fails.

Many have thought that whatever the status of the schematic principle Substitution, the quantified principle Quantified Substitution is unimpeachable. Quine, for example, declares that "violating substitutivity of identity for variables…would simply be a wanton misuse of the identity sign."[6] Kaplan writes that "Quine justly accuses of wantonness" those who reject Quantified Substitution, before mounting his own defense of it.[7] And while some who allow for failures of Substitution have also countenanced failures of Quantified Substitution, typically proponents of this view have nonetheless thought that Leibniz's Law should still be upheld. Thus, Richard says "In suggesting that [Quantified Substitution] is not a truth of quantification theory, I do NOT impugn the trivial truth that things which are one share their properties."[8]

Let *Classical Opacitism* be, roughly, the view that there are cases of opacity, but they should not lead us to reject classical quantification theory. In this paper we will consider

---

2 Throughout this paper we will focus on attitude reports as a motivation for rejecting Substitution. But there are other reasons one might be motivated to reject Substitution, such as puzzles of material constitution (Geach (1967), Lewis (1971), Gibbard (1975), Gupta (1980), but see Fine (2003)), vague identities (for discussion, see Heck (1998), Williamson (2002) and Edgington (2002), responding to Evans (1978)), and, more recently, counterpossibles (Kocurek (forthcoming)). For surveys of these issues, see Hawthorne (2003) and Magidor (2011).

3 Salmon (1986), Soames (1987), Braun (1998), Saul (1997, 2010).

4 Schiffer (1979), Crimmins & Perry (1989), Crimmins (1992), Dorr (2014b), Goodman & Lederman (2018).

5 Fregeans – although not Frege himself (1892) – sometimes claim that they do not reject Substitution, see e.g. Carnap (1947, p. 137) and Kaplan (1968, pp. 183-4). Goodman & Lederman (2019) argue against Carnap and Kaplan that Fregeans are in fact committed to rejecting Substitution.

6 Quine (1986, p. 339) cf. Quine (1943), Quine (1960, p. 151 cf. §35 and §41) and Quine (1961).

7 Kaplan (1986, p. 244) cf. p. 235 and Appendix D.

8 Richard (1987, p. 555 n. 1). Richard only rejects instances of Quantified Substitution in which either $\forall x$ binds more than one occurrence of x in φ or $\forall y$ binds more than one occurrence of y in $\varphi[y/x]$; so, unlike the views we will be exploring, accepting Hesperus/Phosphus would commit him to rejecting classical quantificational logic.

various logics of identity consistent with opacity, and defend Classical Opacitism against objections. We conclude that, far from being indefensible, Classical Opacitism is perhaps the most attractive position compatible with opacity.[9]

Section 2 introduces the framework of higher-order logic within which our investigation will take place, and shows how Substitution may be derived from other principles concerning identity within that framework. This derivation highlights a crucial decision point for any opacity-friendly account of identity. In Section 3, we provide a precise characterization of Classical Opacitism and develop theories of Classical Opacity based on different ways of responding to the choice point described in Section 2. In Sections 4 and 5, we defend these theories against a number of objections. We also consider some non-classical logics of identity that are consistent with opacity, and argue that Classical Opacity compares favorably to them.

Appendix A is a self-contained presentation of the basics of the model theory of higher-order logic. Appendices B and C present a series of consistency results for the main theories of classical opacity developed in Section 3. Appendix D presents models of Free Opacity, a theory developed in Section 4. Appendix E contains a glossary of principles we discuss in the paper. These appendices contain a number of further results not discussed in the main text.

2 Substitution

Our opening examples concerned identities between individuals. But in addition to the identity predicate of first-order logic, which combines with two singular terms to form a sentence, there are also idioms of identification that combine with two sentences, or with two predicates, to form a sentence. In English, for example, the idiom "For it to be the case that ... just is for it to be the case that ..." allows us to form such identifications from a pair of sentences, and "To be ... just is to be ..." allows us to form such identifications given a pair of predicates.[10] For example, we might say: "For it to be the case that Boston is north of New York just is for it to be the case that New York is south of Boston"; or "To be mostly water just is to be mostly H_2O".

In what follows we will understand the phenomenon of opacity broadly, so that it includes failures of such identifications to license the intersubstitution of the relevant sentences and predicates.[11] Those who think that Hesperus/Phosphorus is a case of opacity involving the first-order identity predicate should also think that there are 'higher-order' cases of opacity too, such as:

Propositional Hesperus/Phosphorus: For it to be the case that Hesperus is visible at night just is for it to be the case that Phosphorus is visible at night. But while the ancients knew that Hesperus is visible at night, they did not know that Phosphorus is visible at night.

---

[9] Carnap (1947), Richard (1987) and Priest (2005, Ch. 2) are (in different ways) precedents for such views. For discussion of Carnap see Williamson (2013, §2.4).

[10] For more on higher-order identifications see Rayo (2013), Dorr (2016).

[11] An early argument for the existence of higher-order cases of opacity (though couched in somewhat different terms) can be found in in Mates (1952).

Property Hesperus/Phosphorus: To point to Hesperus just is to point to Phosphorus. But while the ancients knew that at night they could point to Hesperus, they did not know that at night they could point to Phosphorus.

In light of examples of this kind, our investigation of which theories are consistent with opacity will concern not only the logic of identity between individuals, but also the logic of higher-order identifications.

It turns out that there are important connections between principles governing identity at higher types and Substitution. This section will be devoted to an argument which exhibits one such connection.

The argument goes as follows:

(i) Hesperus is Phosphorus.

(ii) So to be instantiated by Hesperus just is to be instantiated by Phosphorus.

(iii) So for *being known by the ancients to be visible at night* to be instantiated by Hesperus just is for *being known by the ancients to be visible at night* to be instantiated by Phosphorus.

(iv) So for the ancients to know that Hesperus is visible at night just is for the ancients to know that Phosphorus is visible at night.

(v) So the ancients knew that Hesperus was visible at night if and only if the ancients knew that Phosphorus was visible at night.

Parallel arguments, if valid, can be used to establish any instance of Substitution. So opacitists must reject the validity of at least one of the argument's four steps. Which step they reject turns out to be one of the most important decision points in developing an opacity-friendly theory of identity.

Assessing arguments like the one above requires a framework for theorizing about properties, properties of properties, and identifications involving them. To this end, we will consider theories formulated in the language of higher-order logic – in particular, the (simply, functionally) typed lambda-calculus. We'll now introduce this language, present the preceding argument formally, and introduce the Core Theory, which will serve as the logical backdrop for the rest of the paper.

In our language, terms are classified into different syntactic categories called *types*. There are two basic types: e, the syntactic category of individual constants (i.e., expressions regimenting proper names like 'Hesperus'), and t, the syntactic category of formulas. All other types are characterized in relation to e and t, as follows.

For any two types σ and $\tau \neq e$, there is a third type $(\sigma \rightarrow \tau)$; a term F is of this type just in case, for every term a of type σ, $\ulcorner Fa \urcorner$ is a term of type τ. Since predicates like 'is visible at night' combine with names like 'Hesperus' (type e) to form sentences like 'Hesperus is visible at night' (type t), they would be regimented as terms of type $(e \rightarrow t)$. Similarly, sentential negation, which combines with a sentence to form another sentence, is regimented as a term '$\neg$' of type $(t \rightarrow t)$. Polyadic predicates are regimented so as to take their arguments one at a time. For instance, 'loves', which takes two

arguments, is regimented so that it combines first with 'Mary' (type e) to yield 'loves Mary', which combines with 'John' (type e) to yield 'John loves Mary' (type t). Since 'loves Mary' is regimented as a term of type $(e \to t)$, 'loves' will be regimented as a term of type $(e \to (e \to t))$ (which we abbreviate $e \to e \to t$, restoring omitted parentheses on the right). Similarly, sentential conjunction '$\wedge$' is regimented as a term of type $t \to t \to t$, since it combines with two sentences (one at a time) to yield a sentence.[12]

For every type σ, our language includes an identification constant $=_\sigma$ of type $\sigma \to \sigma \to t$. Thus, where F stands for 'is mostly water' and G for 'is mostly H_2O', we can regiment the identification 'to be mostly water just is to be mostly H_2O' as $\ulcorner F =_{e \to t} G \urcorner$.

Our language also includes infinitely many variables of every type. If x is a variable of type σ and φ a term of type $\tau \neq e$, then $\ulcorner (\lambda x.\varphi) \urcorner$ is a term of type $\sigma \to \tau$ in which any free occurrences of x in φ are bound by the prefix $\ulcorner \lambda x \urcorner$. In the case where φ is a formula in which x occurs free, we might pronounce this term $\ulcorner$is such that φ['it'/x]$\urcorner$, subject to the convention that the introduced occurrences of 'it' be interpreted as anaphoric on a in any formula $\ulcorner (\lambda x.\varphi) a \urcorner$. For example, $\ulcorner (\lambda x.x$ is visible at night)Hesperus$\urcorner$ might be pronounced 'Hesperus is such that it is visible at night'.

Finally, for every type σ, we have a universal quantifier $\forall_\sigma$ of type $(\sigma \to t) \to t$. These quantifiers combine with predicates to form sentences; e.g., $\ulcorner \forall_e F^{e \to t} \urcorner$, which we might pronounce "everything is F." This can seem odd to those used to first-order logic. In standard first-order languages, quantifiers do double-duty. On the one hand, they bind variables; on the other hand, they are devices for expressing generality. By contrast, in the higher-order language we will use, variable binding is handled solely by λ-terms, and quantifiers are predicate constants rather than variable-binding sentential operators. However, for readability, we will often adopt a more familiar notation, letting $\ulcorner \forall x \varphi \urcorner$ abbreviate $\ulcorner \forall_\sigma (\lambda x.\varphi) \urcorner$, when x is a variable of type σ and φ a term of type t. We will also let $\ulcorner \exists x \varphi \urcorner$ abbreviate $\ulcorner \neg \forall_\sigma (\lambda x.\neg \varphi) \urcorner$.

While type systems like the one just described are sometimes used to theorize about natural languages, the above English glosses serve merely to illustrate the basic idea behind the stipulated formation rules for our formal higher-order language. We are not making any substantive claims about the syntax of English or about the possibility of translating between English and higher-order languages. For ease of pronunciation, we will sometimes speak of 'properties of individuals', 'binary relations between individuals', 'properties of properties of propositions' and the like, but this is merely a strategy for conveying in prose claims that are officially formulated in our higher-order language using terms of types $e \to t$, $e \to e \to t$, $(t \to t) \to t$, and so on.[13] In the introduction when we said "x and y have the same properties", we already meant to be employing this strategy.

We will sometimes indicate the type of a term with a superscript on its first occurrence (as we did above with $F^{e \to t}$). When we omit type indications, this is either because they can be inferred from context, or because the term in question is a schematic letter in a 'type-

[12] For readability, we will typically use infix notation for $\wedge$ and other binary connectives, writing $\ulcorner \varphi \wedge \psi \urcorner$ instead of $\ulcorner \wedge \varphi \psi \urcorner$. We will also freely include/omit parentheses around predications.

[13] This attitude towards theorizing in higher-order languages is similar to that advocated by Prior (1971), Williamson (2003) and Dorr (2016).

ambiguous' schema whose instances should be understood as including all uniform substitutions of terms for schematic letters that result in a well-formed formula. Substitution, Quantified Substitution and Leibniz's Law should be understood as type-ambiguous schemas of this sort.

We can now formalize the argument from the beginning of this section by appeal to the following principles:

Lift Congruence $a = b \rightarrow (\lambda X.Xa) = (\lambda X.Xb)$,

Application Congruence $F = G \rightarrow Fa = Ga$,

Beta-Eta Equivalence $\varphi \leftrightarrow \psi$, provided φ and ψ are $\beta\eta$-equivalent,[14]

Material Equivalence $p = q \rightarrow (p \leftrightarrow q)$.

To establish a given instance of Substitution, suppose (i) $a = b$. Then (ii) $(\lambda X.Xa) = (\lambda X.Xb)$ (by Lift Congruence), so (iii) $(\lambda X.Xa)(\lambda x.\varphi[x/a]) = (\lambda X.Xb)(\lambda x.\varphi[x/a])$ (by Application Congruence). Thus, we have (iv) $(\lambda x.\varphi[x/a])a = (\lambda x.\varphi[x/a])b$, and hence $\varphi = \varphi[b/a]$ (by Beta-Eta Equivalence). And so, finally, we have (v) $\varphi \leftrightarrow \varphi[b/a]$ (by Material Equivalence).

In what follows we will explore views that reject either Lift Congruence or Application Congruence. We have little to say in support of Material Equivalence other than that it strikes us as unimpeachable.

Beta-Eta Equivalence is much more controversial. Here is not the place to undertake a general defense of the principle, other than to note that it is an orthodox principle of higher-order logic.[15] But we do want to make two more local observations about the role of the principle in the above derivation of Substitution.

One common reservation about Beta-Eta Equivalence is that it is incompatible with the view that predications are 'structured' in the manner of the sentences that express them, so that $Fa = Gb$ only if $F = G$ and $a = b$. For instance, Beta-Eta Equivalence implies that $(\lambda x$.John loves $x)$Mary = $(\lambda x.x$ loves Mary)John, despite the fact that $(\lambda x$.John loves $x) \neq (\lambda x.x$ loves Mary) and John $\neq$ Mary. However, this sort of reservation about Beta-Eta Equivalence is orthogonal to the questions about opacity that we will be considering here. This is because the above derivation of Substitution does not require the full strength of Beta-Eta Equivalence, but only the principle:

Reduction Congruence $(\lambda x.\varphi)a = (\lambda x.\psi)b \rightarrow \varphi[a/x] = \psi[b/x]$.

Unlike Beta-Eta Equivalence, there is no obvious tension between Reduction Congruence and the picture of propositions as being structured like sentences, and we expect that most philosophers attracted to that picture will find Reduction Congruence congenial. In proofs in what follows we'll cite Reduction Congruence where possible to illustrate

[14] Two terms are $\beta\eta$-equivalent just in case one can be obtained from the other by a sequence of substitutions of sub-terms that are either *immediately η-equivalent* – i.e., of the form F and $\ulcorner(\lambda x.Fx)\urcorner$, with x not free in F – or *immediately β-equivalent* – i.e., of the form $\varphi[a/x]$ and $\ulcorner(\lambda x.\varphi)a\urcorner$, with a free for x in φ.

[15] For further discussion of Beta-Eta Equivalence see Dorr (2016, §5-6).

which arguments should be acceptable to those who believe propositions are 'structured' in this way.[16]

A second reservation about Beta-Eta Equivalence is explicitly tied up with issues related to opacity. The idea is that even those who think that Hesperus/Phosphorus is a case of opacity should accept the schema:

e-**Atomic Substitution** $a^e = b^e \rightarrow (Fa \leftrightarrow Fb)$.

Informally, the idea is that there is no opacity involving atomic predications of individuals.[17] Those who think that Hesperus/Phosphorus is a case of opacity but wish to preserve *e*-Atomic Substitution can reject the inference from (iii) to (iv) in the above argument for Substitution, and hence reject Beta-Eta Equivalence. Having done so, they might think that rejecting Beta-Eta Equivalence should allow friends of opacity to accept both Lift Congruence and Application Congruence, and thereby sidestep a central project of the present paper.

We think that this strategy is much less attractive than it might appear at first blush, for reasons that only emerge when we consider higher-order identifications. The reason is that, if Hesperus/Phosphorus presents a counterexample to Substitution, then Propositional Hesperus/Phosphorus presents a counterexample to:

t-**Atomic Substitution** $a^t = b^t \rightarrow (Fa \leftrightarrow Fb)$.

Our own view is that *e*-Atomic Substitution and *t*-Atomic Substitution should stand or fall together. But even those who do not share this view should note that *t*-Atomic Substitution can be derived from Lift Congruence and Application Congruence using only the following extremely weak consequence of Beta-Eta Equivalence:

Instantiation Equivalence $(\lambda X.Xa)F \leftrightarrow Fa$.[18]

Thus those who want to maintain Lift Congruence and Application Congruence while accepting Propositional Hesperus/Phosphorus (and so rejecting *t*-Atomic Substitution) will have to reject Beta-Eta Equivalence in a particularly radical and thoroughgoing way. We don't think that such views are uninteresting, and we hope to explore them in further work. But they are sufficiently radical and disruptive to compelling principles involving

---

[16] Note also that Reduction Congruence does not in any sense involve 'quantifying in', since our formal language separates the job of variable binding from the job of quantification. We cannot take for granted that $\ulcorner(\lambda x.\varphi)\urcorner$ can be faithfully glossed as $\ulcorner$is *an* x such that $\varphi\urcorner$ – i.e., that it is equivalent to $\ulcorner(\lambda y.\exists x(x = y \wedge \varphi))\urcorner$. (Goodman (2016) makes a parallel point in reply to those like like Stalnaker (2012) who maintain that $(\lambda x.(\lambda y.\neg\exists z(z = y))x)a = (\lambda x.\neg\exists z(z = x))a \rightarrow (\lambda y.\neg\exists z(z = y))a = \neg\exists z$ $(z = a)$ is a counterexample to Reduction Congruence for a that only contingently exist.) This is important to keep in mind, since below we will consider the view of Bacon & Russell (2017) according to which not only does this equivalence fail, but the above argument would fail only at step (iii) were the λ-terms involved replaced with their quantification-involving counterparts.

[17] This idea is often run together with idea that Quantified Substitution is unimpeachable, under the slogan that there can be no opacity '*de re*'; cf. Quine (1956) and Kaplan (1968) with Cumming (2016).

[18] The proof goes as follows: suppose (i) $p = q$. Then (ii) $(\lambda X.Xp) = (\lambda X.Xq)$ (Lift Conguence); (iii) so, $(\lambda X.Xq)B = (\lambda X.Xp)B$ (Application Congruence); (iv) so, $(\lambda X.Xp)B \leftrightarrow (\lambda X.Xq)B$ (Material Equivalence); (v) so, $Bp \leftrightarrow Bq$ (Instantiation Equivalence).

predicate abstraction (e.g. Instantiation Equivalence) to motivate the more conservative strategy we adopt in this paper.

Let a *theory* be a set of sentences T of our higher-order language that: (i) contains every propositional tautology; (ii) is closed under Modus Ponens (so that if $\varphi \rightarrow \psi \in T$ and $\varphi \in T$ then $\psi \in T$) ; and (iii) satisfies the following closure condition:

Generalization If $\varphi \rightarrow \psi \in T$, then $\varphi \rightarrow \forall x\psi \in T$, where x does not occur free in φ.

We'll often call the elements of a theory its *theorems*.[19] We'll say that a theory is *consistent with opacity* just in case some instances of Substitution are not theorems of it.

Call the theory axiomatized by

Equivalence $a = a \wedge (a = b \wedge a = c \rightarrow b = c)$,

Material Equivalence, and Beta-Eta Equivalence the *Core Theory*.[20] Our aim is to study extensions of the Core Theory which are consistent with opacity. No such extension can include both Lift Congruence and Application Congruence, since, as we have seen, they jointly entail Substitution given the Core Theory.[21] Thus the theories that we will consider will either not contain every instance of Lift Congruence or not contain every instance of Application Congruence.

3 Classical Opacity

The Core Theory does not contain standard axioms governing the quantifiers. The remainder of the paper will study extensions of the Core Theory; a key question will be which quantificational axioms should be included in those extensions. The simplest, best-known extension of the Core Theory with quantificational axioms is what we will call *classical higher-order logic*, the theory axiomatized by the Core Theory and

Universal Instantiation $\forall x\varphi \rightarrow \varphi[a//x]$, where a is free for x in φ.[22]

We emphasize that, as we will understand it, classical higher-order logic is consistent with opacity; it does not contain every instance of Substitution. It stands in contrast to *transparent higher-order logic*, the theory axiomatized by Universal Instantiation, Substitution and the Core Theory.

Classical higher-order logic has many virtues. Notice that, Universal Instantiation is equivalent to Existential Introduction: $\varphi[a//x] \rightarrow \exists x\varphi$. In the first-order setting this

---

19 A schema is a theorem of a theory if every instance of the schema is a theorem of the theory.

20 A theory is axiomatized by some schemas just in case it is the smallest theory containing every instance of those schemas. We say that φ *follows from* Γ in T or that Γ *entails* φ or that T and Γ entails φ whenever there are some $\psi_0, \ldots, \psi_n$ in Γ so that $\top \wedge \psi_0 \wedge \ldots \wedge \psi_n \rightarrow \varphi \in T$, where $\top$ is some tautology. Note that, given Equivalence, Beta-Eta Equivalence entails that if φ and ψ are $\beta\eta$-equivalent, then $\varphi = \psi$.

21 It's worth noting that given the Core Theory, Substitution also entails Lift Congruence and Application Congruence. Thus, given the Core Theory, Substitution is equivalent to the conjunction of Lift Congruence and Application Congruence.

22 Here $\varphi[a//x]$ is obtained from φ by replacing every free occurrence of x in φ with an occurrence of a. A term a is free for x in φ just in case no free occurrence of a variable in a becomes bound when a is substituted for a free occurrence of x in φ.

schema provides a simple, natural explanation of why we can infer "There's a planet the ancients knew was visible at night" from "Hesperus is a planet and the ancients knew Hesperus was visible at night". In our higher-order setting it also provides a simple, natural explanation of why we can infer "There's something Hesperus was, namely, known by the ancients to be visible at night" from "Hesperus was known by the ancients to be visible at night." Classical higher-order logic is therefore a natural and conservative starting point in developing theories of identity consistent with opacity. We call such theories theories of *Classical Opacity*, and their proponents *Classical Opacitists*.

Given the Core Theory, Universal Instantiation implies that if there are cases of opacity, then Leibniz's Law has false instances. In particular, if $a = b \rightarrow (\varphi \leftrightarrow \varphi[b/a])$ is a false instance of Substitution, then $a = b \rightarrow \forall X(Xa \leftrightarrow Xb)$ is a false instance of Leibniz's Law.[23]

Say that a and b are *Leibniz equivalent*, written $a \approx b$, just in case they have all the same properties. Leibniz equivalence can be formally defined as follows: $\approx =_{df} (\lambda x^{\sigma} \lambda y^{\sigma}.\forall X^{\sigma \rightarrow t}(Xx \leftrightarrow Xy))$.[24] Classical Opacitists hold that cases of opacity are cases in which some x and y are identical but not Leibniz equivalent. They thus reject the apparent truism with which we began.

In this section we will develop theories of Classical Opacity. We first consider theories which include Application Congruence (and hence do not include Lift Congruence). We then consider theories which include Lift Congruence (and hence do not include Application Congruence). In later sections we defend Classical Opacity by responding to objections to it and arguing that it compares favorably to other extensions of the Core Theory consistent with opacity.

3.1 Applicativism

The first theory of classical opacity upholds Application Congruence.

A minimal principle motivating this view is that if to be F is to be G then anything which is F is G, i.e. that $F = G \rightarrow \forall x(Fx \leftrightarrow Gx)$. Put simply, identical properties must have the same extension. Given Universal Instantiation, the claim that identical properties have the same extension implies:

Application Equivalence $F = G \rightarrow (Fa \leftrightarrow Ga)$.

Application Equivalence is weaker than Application Congruence against the background of the Core Theory; the former follows from the latter using Material Equivalence, but there is no entailment in the other direction.

Earlier we showed that Application Congruence and Lift Congruence together imply Substitution given the Core Theory. But in fact, we can show that Application

---

[23] The implication follows by propositional logic given $\varphi \leftrightarrow (\lambda x.\varphi[x/a])a$ and $\varphi[b/a] \leftrightarrow (\lambda x.\varphi[x/a])b$, which are instances of Beta-Eta Equivalence, and $\forall X(a = b \rightarrow (Xa \leftrightarrow Xb)) \rightarrow (a = b \rightarrow ((\lambda x.\varphi[x/a])a \leftrightarrow (\lambda x.\varphi[x/a])b))$, which is an instance of Universal Instantiation – where x is any variable occurring in neither φ nor b.

A parallel argument not requiring Beta-Eta Equivalence shows that false instances of t-Atomic Substitution (e.g., Propositional Hesperus/Phosphorus) are incompatible with Leibniz's Law given Universal Instantiation, and indeed, even given a restriction of Universal Instantiation to 'atomic' predications, i.e. $\forall xFx \rightarrow Fa$.

[24] As in the case of conjunction and identity, we typically use infix notation for this defined relation.

Equivalence together with Lift Congruence is sufficient to derive Substitution against the background of the Core Theory.[25] So Classical Opacitists who accept the idea that identical properties have the same extension must reject Lift Congruence. And once Lift Congruence is rejected, it is natural for those attracted to Application Equivalence to accept Application Congruence too.

Since theories are closed under Generalization, any theory that contains Application Congruence also contains:

Quantified Application Congruence $F = G \rightarrow \forall x(Fx = Gx)$.

And any theory that contains Quantified Application Congruence and Universal Instantiation must also contain Application Congruence. So Application Congruence and Quantified Application Congruence are equivalent given Classical Opacity.

Let *Functionality* be the converse of Quantified Application Congruence, i.e. $\forall x(Fx = Gx) \rightarrow F = G$. In this section we will focus on the following biconditional, which is equivalent to the conjunction of Functionality and Quantified Application Congruence:

Applicative Individuation $\forall x(Fx = Gx) \leftrightarrow F = G$.

The theory that we will be developing is primarily motivated by an interest in the combination of Classical Opacity and Application Congruence. However, since we can show that a stronger theory which also includes Functionality is consistent, we will assume Functionality in our discussion as well.[26]

Given Applicative Individuation, once the extension of the identity relation is fixed at the base types t and e, it is fixed for all other types. More specifically, given Applicative Individuation, for any σ and τ, $=_{\sigma\rightarrow\tau}$ is determined by $=_{\tau}$. For, given Applicative Individuation, we have: $F =_{\sigma\rightarrow\tau} G \leftrightarrow \forall x^{\sigma}(Fx =_{\tau} Gx)$.

Even if there are some cases of opacity, a reasonable theory of identity should entail that there is a large class of *transparent* expressions, where an expression θ is transparent in a theory T just in case the schema $a = b \rightarrow \theta a = \theta b$ is a theorem of T.[27] In particular, it is natural to think that broadly logical vocabulary such as quantifiers, Boolean operators and identity should be transparent. The transparency of such logical vocabulary is not entailed by anything we have said so far, so we would like to consider further principles which do entail their transparency.

Indeed, there is a more general class of terms whose transparency we want to ensure. We can inductively define Boolean operators for each type $\sigma \rightarrow \tau$, letting $\neg^{\sigma\rightarrow\tau} =_{df} \lambda f^{\sigma\rightarrow\tau}\lambda x.\neg^{\tau} fx$, and $\wedge^{\sigma\rightarrow\tau} =_{df} \lambda f^{\sigma\rightarrow\tau}\lambda g^{\sigma\rightarrow\tau}\lambda x.fx \wedge^{\tau} gx$. Call the quantifiers, identity connectives, ordinary Boolean operators and these extensions of the Boolean operators *extended logical constants*.

---

[25] Indeed, Lift Congruence and Application Equivalence are equivalent to Substitution against that background.

[26] Functionality is independent of the rest of the theory we develop here; see Benzmüller et al. (2004). On the metaphysical motivation for higher-order logics in which Functionality fails, see Fritz & Goodman (2016) and Bacon (2018).

[27] Given Material Equivalence, if an expression of type $\sigma \rightarrow t$ is transparent in this sense, then it will also be transparent in the more familiar sense of obeying the schema: $a =_{\sigma} b \rightarrow (\theta a \leftrightarrow \theta b)$.

We then impose the following axioms:

Constant Transparency $a = b \rightarrow Oa = Ob$, if O is an extended logical constant.

Transparency Preservation $a = b \rightarrow (Oc)a = (Oc)b$, if O is an extended logical constant.

Constant Transparency implies, for example, that if $p = q$ then $\neg p = \neg q$ and if $F = G$ then $\forall F = \forall G$. It also implies that $p = q \rightarrow (\wedge p) = (\wedge q)$. But Constant Transparency does not on its own imply that $p = q \rightarrow (\wedge r)p = (\wedge r)q$: the operator $(\wedge r)$ is not an extended logical constant, so it is consistent with Constant Transparency that this new operator fail to be transparent. Transparency Preservation rules out this possibility, by requiring generally that operators produced by applying binary extended logical constants should also be transparent.[28] In the presence of Application Congruence, these two principles together eliminate the odd behavior which was consistent with Constant Transparency alone. They imply that identity is a congruence both with respect to the logical constants $=_\sigma$ and the extended logical constants $\wedge^\sigma$, in the following sense:

Identity Congruence $(a = b \wedge c = d) \rightarrow (a = c) = (b = d)$

Conjunction Congruence $(a = b \wedge c = d) \rightarrow (a \wedge c) = (b \wedge d)$.[29]

Let *Classical Applicativism* – so called because it upholds Application Congruence – be the theory axiomatized by Applicative Individuation, Constant Transparency, Transparency Preservation and classical higher-order logic.

As will become evident in the course of the paper, it is not obvious that theories of the kind we are considering are consistent with opacity. Innocuous-seeming assumptions often turn out to entail Substitution. In Appendix B, though, we show that Classical Applicativism is consistent with opacity by providing a model in which there is opacity but all of the principles of Classical Applicativism hold.

It's worth emphasizing that the models that we appeal to are merely tools for demonstrating the consistency of the views we are interested in. We don't take these models to be probative regarding the question of how to develop a theory of meaning for our formal language on its intended interpretation.[30] Generally, the models we give for theories of Classical Opacity invalidate instances of Substitution by allowing $\langle x, y \rangle$ to be in the extension of an identity connective despite x and y being numerically distinct objects in the domain of the model. Deploying this strategy for consistency proofs should not be confused with claiming (absurdly) that "=" does not mean identity, or that Hesperus isn't

---

[28] More generally, in a language enriched with $n + m$-ary connectives as logical constants, we should require that saturating the first n of its arguments will yield a transparent predicate (and likewise for higher-type 'extensions' of the connective):

Strong Transparency Preservation $a = b \rightarrow (((Oc_1)\ldots c_n)a = ((Oc_1)\ldots c_n)b)$, for O an extended logical constant of the enriched language.

[29] By Constant Transparency, assuming $a = b$, we have $(= a) = (= b)$. By Application Congruence we have $(a = c) = (b = c)$. By Transparency Preservation, assuming $c = d$, we have $(b = c) = (b = d)$. And so, by Equivalence, we have $(a = c) = (b = d)$. The argument is essentially the same in the case of $\wedge^\sigma$.

[30] For more on this distinction between model theory and semantics see see Burgess (2008).

really Phosphorus.[31] The question of how to give a semantic theory for a language exhibiting opacity is a rich and important one which we hope to explore in future work, but it is not the topic of this paper.

Just as there are apparent cases of opacity concerning individuals and properties, so too are there apparent cases of opacity concerning propositions. We saw one case above in Propositional Hesperus/Phosphorus. But there are putative cases of propositional opacity that are not in any obvious way attributable to first-order opacity. Here are three such examples:

Absorption Identification: For it to be the case that Athens is a city is for it to be the case that both Athens is a city and either Athens is a city or there are neutrinos. But while the Greeks knew that Athens is a city, they did not know that both Athens is a city and either Athens is a city or there are neutrinos.

Intuitionism: For Harry to be not not either bald or not bald just is for Harry to be either bald or not bald. But while Crispin believes that Harry is not not either bald or not bald, Crispin does not believe that Harry is either bald or not bald.

Mathematics: For it to be the case that it is raining if it is raining is for it to be the case that if the axioms of second order arithmetic hold, then Fermat's Last Theorem holds. But while Euler knew that it is raining if it is raining, Euler did not know that if the axioms of second order arithmetic hold, then Fermat's Last Theorem holds.

The first sentences of each of these examples are controversial claims about propositional fineness of grain. Depending on one's theory of propositional identity, putative cases of opacity like the ones above may be more or less widespread. In particular, coarse-grained theories of propositions – which entail many such identifications – generate many putative examples of opacity of this kind.

We've already claimed in Section 2 that if one accepts Hesperus/Phosphorus, then one should accept Propositional Hesperus/Phosphorus too. Our question now is whether Classical Opacitists can treat the above putative examples of propositional opacity in the same way as they treat Propositional Hesperus/Phosphorus. That is, can the Classical Opacitist accept a coarse-grained theory of propositional identity, while holding that Leibniz-equivalence draws much finer distinctions? If they can, they would be able to reconcile coarse-grained theories of propositions with the possibility of cognitive accomplishment in logic, i.e. to resolve (one version of) the so-called "problem of logical omniscience".

It is an interesting question how far this thought can be taken. Can the Classical Opacitist endorse a *very* coarse grained theory of propositions, while nevertheless doing justice to claims like the ones above about differing cognitive significance, by claiming that

[31] Ripley (2012) also develops models in which 'Hesperus' and 'Phosphorus' denote different objects. But Ripley takes there to be an intended model of this form, and suggests that strictly Hesperus and Phosphorus are not identical. We take it to be obvious that Hesperus and Phosphorus are identical, though we are exploring views on which Hesperus and Phosphorus have different properties.

the propositions in question fail to be Leibniz equivalent? A simple and familiar coarse-grained theory of propositions is:

Booleanism $a = b$, whenever $a \leftrightarrow b$ is a theorem of propositional logic.[32]

The models of Classical Applicativism developed in Appendix B show that the Classical Opacitist can endorse Booleanism while holding that propositions are quite fine-grained at the level of Leibniz equivalence. The idea that there is no distinction at the level of identity between the propositions expressed by tautologically equivalent sentences is an attractive one. For example, it is not implausible that for it to be the case that φ just is for it to be the case that $(\varphi \wedge \psi) \vee (\varphi \wedge \neg\psi)$. Our models show that the Classical Opacitist can accept such identifications, while also doing justice to the possibility that logically equivalent sentences can express propositions which differ in cognitive significance. Given such differences in granularity at the levels of identity and Leibniz equivalence, the Classical Opacitist can see failures of logical omniscience as continuous with more familiar apparent cases of opacity that arise concerning individuals such as Hesperus and Phosporus.[33]

3.2 Purity

Constant Transparency was motivated by the idea that logical vocabulary should be transparent. The logical constants, however, do not exhaust the logical vocabulary of our higher-order language. In particular, closed terms containing no constants at all (i.e. formed by λ-abstraction on formulas featuring only variables) are naturally understood to be logical vocabulary as well. Such terms are called *combinators*.

Importantly, Classical Applicativism together with the transparency of combinators would rule out opacity. For the transparency of any combinator of the form $\lambda y \lambda X.Xy$ implies Lift Congruence, and as we have seen Application Congruence and Lift Congruence jointly imply Substitution.[34] Thus the Classical Applicativist must reject the claim that such combinators are transparent.

In this section, we'll consider theories of Classical Opacity that hold on to the idea that all combinators are transparent, and accordingly reject Application Congruence.

We'll say that a *pure term* is any closed expression that contains no free variables and no constants other than the logical constants $\neg, \wedge, =$, and $\forall$. The first theory that we'll consider holds that all pure terms are transparent:

---

[32] See Dorr (2016, §7) for discussion of Booleanism and related views which strengthen Booleanism to include analogous identifications at other types as well.

[33] There are, however, important limits to how far this strategy can be taken. In particular, Classical Opacitists cannot hold that, when it comes to Leibniz equivalence, propositions are structured in the manner of the sentences we use to express them – at least, not on a naïve way of articulating that idea. This is because $\neg\forall X^{\iota\to\iota}\forall Y^{\iota\to\iota}(X\varphi \approx Y\varphi \to X \approx Y)$ is a theorem of classical higher-order logic. In other words, it is false that that predications are only Leibniz equivalent when they are predications of Leibniz-equivalent properties. Note that the above formula is *not* a theorem of the Core Theory, and there are indeed strategies for non-classical opacitists to uphold the idea that propositions are structured naively in the manner of sentences. For proof and discussion of these results, see Goodman (2017).

[34] If $a = b$, then $(\lambda y \lambda X.Xy)a = (\lambda y \lambda X.Xy)b$, by the transparency of $(\lambda y \lambda X.Xy)$, and hence $(\lambda X.Xa) = (\lambda X.Xb)$, by Reduction Congruence.

Purity $a = b \rightarrow Fa = Fb$, provided F is a pure term.[35]

Purity guarantees, in a simple and principled way, the existence of a large class of transparent properties. Like Constant Transparency, Purity guarantees that the Boolean operators, quantifiers and identity are all transparent. But Purity does much more than this: it also implies that any expressions built up from these basic operators and λ-abstraction are also transparent.

Purity, however, does not entail Transparency Preservation, and more importantly, it does not entail Identity Congruence or Conjunction Congruence. The reason is that pure terms may be corrupted by application. For instance, although $\wedge$ is a pure term, if r is not pure, $(\wedge r)$ is not pure. So while Purity entails that $\wedge$ is transparent, it does not entail that $(\wedge r)$ is transparent. Similar reasoning establishes that while Purity entails that $=$ is transparent, it does not entail that $(=r)$ is transparent if r is not pure.

It is not just that Purity fails to entail these claims. The Core Theory, Purity, Conjunction Congruence and Universal Instantiation together entail Substitution.[36] A Classical Opacitist who endorses Purity must therefore reject the claim that identity is a congruence with respect to all extended logical constants.

Despite this result, the Classical Opacitist who endorses Purity can still endorse Identity Congruence, the claim that identity is a congruence with respect to identity. They can also endorse the claim that identity is a congruence with respect to *propositional* conjunction:

Propositional Conjunction Congruence $(p =_t p' \wedge q =_t q') \rightarrow (p \wedge q =_t p' \wedge q')$.

Let *Classical Purity* be the theory axiomatized by Propositional Conjunction Congruence, Identity Congruence, Purity and classical higher-order logic. In Appendix C.2, we prove that Classical Purity is consistent with opacity.

Classical Purity is a strong, simple theory. But as we have already seen in the case of Conjunction Congruence, its strength can lead to surprising limitative results. We will now discuss two further such results.

The results both flow from the following lemma, which states that identities involving pure terms imply the corresponding Leibniz equivalences:

Lemma 1: *If a is a pure term, then Classical Purity entails* $a = b \rightarrow a \approx b$.[37]

The first result applies this lemma to exhibit problematic consequences of the conjunction of Classical Purity and Booleanism:

---

[35] We here take inspiration from Bacon & Russell (2017). Their notion of a "Pure Term" differs from ours in that they allow open terms to count as pure. With this qualification, our principle Purity corresponds to their "Pure Transparency". Allowing open terms would render Purity inconsistent with Classical Opacity. For, since theories are closed under Generalization, Leibniz's Law is a theorem of any theory in which variables are transparent.

[36] *Proof:* Suppose $a = b$, Fa and $\neg Fb$. By Purity and Reduction Congruence, $(\lambda X.\neg Xa) = (\lambda X.\neg Xb)$. And so, given Conjunction Congruence, $(\lambda X.Xa \wedge \neg Xa) = (\lambda X.Xa \wedge \neg Xb)$. By Purity, $\exists X(Xa \wedge \neg Xa) = \exists X(Xa \wedge \neg Xb)$, which contradicts Material Equivalence, since $\exists X(Xa \wedge \neg Xb)$, but $\neg\exists X(Xa \wedge \neg Xa)$.

[37] *Proof:* Given Purity and the fact that $(\lambda x \lambda X.Xx \leftrightarrow Xa)$ is a pure term, we have (i) $a = b \rightarrow (\lambda x \lambda X.Xx \leftrightarrow Xa)a = (\lambda x \lambda X.Xx \leftrightarrow Xa)b$. It follows from this, given Reduction Congruence, that we have (ii) $(\lambda X.Xa \leftrightarrow Xa) = (\lambda X.Xb \leftrightarrow Xa)$. And, given this and Purity, we have (iii) $a \approx a = b \approx a$. By Material Equivalence, this implies that (iv) $a \approx a \rightarrow b \approx a$, which, given that $a \approx a$ is a theorem, implies (v) $a \approx b$.

Proposition 2: *Classical Purity and Booleanism entail* $p =_t q \rightarrow (p \rightarrow p) \approx (p \rightarrow q)$.[38]

The idea of the proof is simple. Booleanism implies that any pure propositional tautology $\top$ is such that $\top = (p \rightarrow p)$ for any p. Classical Purity allows us to show that if $p =_t q$ then $(p \rightarrow p) = (p \rightarrow q)$, and hence that $\top = (p \rightarrow q)$. But then by Lemma 1, $(p \rightarrow p) \approx \top$ and $\top \approx (p \rightarrow q)$, so $(p \rightarrow p) \approx (p \rightarrow q)$.

The validity of this conditional is highly problematic given the picture of opacity that we're exploring. On this picture, although for Hesperus to be visible at night just is for Phosphorus to be visible at night, the ancients knew that Hesperus is visible at night, but did not know that Phosphorus is visible at night. Given this basic thought, it is natural to hold further that while the ancients knew that Hesperus is visible at night if Hesperus is visible at night, they did not know that Phosphorus is visible at night if Hesperus is visible at night. But this pattern of judgments is precluded by the combination of Classical Purity and Booleanism. For given Proposition 2, the proposition that Hesperus is visible at night if Hesperus is visible at night is Leibniz equivalent to the proposition that Phosphorus is visible at night if Hesperus is visible at night. Given Universal Instantiation, it follows that the ancients knew that Hesperus is visible at night if Hesperus is visible at night if and only if the ancients knew that Phosphorus is visible at night if Hesperus is visible at night.

Proposition 2 does not show that Booleanism and Classical Purity are inconsistent.[39] But it does show that together they are inconsistent with the sort of judgments which motivate an opacity-friendly account of identity. Proponents of Classical Purity should therefore reject Booleanism.

Classical Purity is also in tension with other assumptions about fineness of grain. Consider for instance:

Identifying Identities $(a = a) = (b = b)$

Identifying Identities is in tension with Classical Purity in the same way that Booleanism is:

Proposition 3: *Classical Purity and Identifying Identities entail* $a = b \rightarrow (a = b) \approx (a = a)$.[40]

---

[38] *Proof:* To show that $p =_t q \rightarrow ((p \rightarrow p) \approx (p \rightarrow q))$, we first show: $p =_t q \rightarrow (p \rightarrow p) = (p \rightarrow q)$. To see that this holds, assume that we have $p =_t q$. Given Purity, we have (i) $\neg p = \neg q$. From this, it follows given Propositional Conjunction Congruence that we have: (ii) $(p \wedge \neg p) = (p \wedge \neg q)$, and so finally we have (iii) $\neg(p \wedge \neg p) = \neg(p \wedge \neg q)$, which, given Booleanism, entails: $(p \rightarrow p) = (p \rightarrow q)$.

Now let $\top$ be $(=_e =_{e \rightarrow e \rightarrow t} =_e) \vee \neg(=_e =_{e \rightarrow e \rightarrow t} =_e)$. Given Booleanism, we have: $\top = (p \rightarrow p)$. And so, given Lemma 1, we have $\top \approx (p \rightarrow p)$. Thus we have, $p =_t q \rightarrow (\top = (p \rightarrow q))$. And so, again given Lemma 1 we have $p =_t q \rightarrow (\top \approx (p \rightarrow q))$. And so, finally, we have, $p =_t q \rightarrow ((p \rightarrow p) \approx (p \rightarrow q))$.

[39] Indeed we have shown that they are consistent in a proof on file with the authors.

[40] *Proof:* Assume $a = b$. Then by Classical Purity we have (i) $(a = b) = (a = a)$ (an instance of Identity Congruence). And by Identifying Identities, we have (ii) $(a = a) = (= = =)$. Given (ii), it follows from Classical Purity (Lemma 1), that we have (iii) $(a = a) \approx (= = =)$. Given this and (i), it follows that we have (iv) $(a = b) = (= = =)$. And so, it follows from (Lemma 1), that we have (v) $(a = b) \approx (= = =)$. It then follows from (iii) that we have (vi) $(a = b) \approx (a = a)$.

The idea of the proof is much the same. Identifying Identities implies that for any pure term π, $(\pi = \pi) = (a = a)$. Classical Purity implies that $(a = a) = (a = b)$ (an instance of Identity Congruence) and hence that $(\pi = \pi) = (a = b)$. But then by Lemma 1, $(a = a) \approx (\pi = \pi)$ and $(\pi = \pi) \approx (a = b)$, and hence $(a = a) \approx (a = b)$.

Once again, such conditionals are in tension with the judgments that motivate an opacity-friendly account of identity. If one accepts that Hesperus was known by the ancients to be visible at night, but Phosphorus was not, it is extremely natural also to accept that the ancients knew that Hesperus is Hesperus, but did not know that Hesperus is Phosphorus. Classical Purity and Identifying Identities preclude this verdict. The proponent of Classical Purity should therefore reject Identifying Identities.

A final corollary of Lemma 1 is that Classical Purists cannot treat cognitive accomplishment in pure logic as a case of opacity (in the way we suggested in the previous section that Classical Applicativists might try to do) since, in the case of pure propositions, any difference in cognitive significance implies numerical distinctness.

Our main models of Classical Purity in Appendix C.2 do not validate Booleanism, Identifying Identities, or the conditionals in Propositions 2 and 3. They are however consistent with independently motivated coarse-grained theories of propositions.[41]

Are there principled theories of identity consistent with opacity that entail the transparency of combinators (and so don't validate Applicative Equivalence) and Booleanism (and so don't validate Purity)? We conclude this section by considering one such theory.

A *pristine term* is any closed term containing at most the constants $\neg$, $\wedge$, and $=$. Every pristine term is pure, but not every pure term is pristine. Proponents of the transparency of combinators (and hence of Lift Congruence) may wish to endorse the following principle, which is weaker than Purity:

Pristineness $a = b \rightarrow Fa = Fb$, provided F is a pristine term.

Let *Classical Pristineness* be the theory axiomatized by Identity Congruence, Propositional Conjunction Congruence, Pristineness, and classical higher-order logic. In Appendix C.1, we provide a model of Classical Pristineness in which Booleanism and Identifying Identities hold, and in which the problematic conditionals mentioned in Propositions 2 and 3 are not validated. It is an open question whether Classical Pristineness and Conjunction Congruence are consistent with opacity.

Propositions 2 and 3 can be proven using only Pristineness and Quantifier Transparency, i.e. $F = G \rightarrow \forall F = \forall G$. This fact illustrates that if one is motivated to reject Purity by the desire to preserve Booleanism (for example), one cannot endorse Constant Transparency; the quantifiers themselves must fail to be transparent.

Classical Applicativism and Classical Purity respond in different ways to the basic result presented in Section 2. Classical Applicativism holds on to Application Congruence and rejects Lift Congruence, while Classical Purity (and Classical Pristineness) holds on to Lift Congruence and rejects Application Congruence. There are at least three significant contrasts between these different theories. (1) Within Classical Purity the transparency of the unary logical operators $\neg$ and $\forall$ arises from a general principle about the transparency of all logical expressions. In Classical Applicativism, by contrast, the transparency of these

[41] See Goodman (forthcoming).

operators is achieved by a comparatively *ad hoc* stipulation. (2) Classical Applicativism is consistent with opacity in the presence of Conjunction Congruence – the principle that identity is a congruence not merely with respect to propositional conjunction, but with respect to conjunction operations at all types. But Conjunction Congruence and Classical Purity are not consistent with opacity. (3) The motivations for opacity fit well with Classical Applicativism given a wide array of theories of coarse-grained propositions. These motivations, however, are in tension with Classical Purity given certain coarse-grained theories of propositions. Proponents of Classical Pristineness can avoid this tension, but only by giving up the principle – endorsed by Classical Applicativists and Classical Purists alike – that if $F = G$, then $\forall F = \forall G$.

4 Leibniz's Law

In the previous section we showed that one can develop strong and simple theories of identity consistent with opacity without rejecting classical quantification theory. In this section, we will consider three arguments against such theories of Classical Opacity. The first two are direct arguments against these theories. The third argument is abductive, concluding that such theories are less attractive than alternative theories consistent with opacity which validate Leibniz's Law. In this connection will develop two such alternatives and argue that they are not in fact preferable to theories of Classical Opacity.

Here is the first argument. If there is opacity, then identity does not satisfy Substitution. But Classical Opacity entails that Leibniz equivalence does satisfy the parallel principle:

LE Substitution $a \approx b \rightarrow (\varphi \leftrightarrow \varphi[b/a])$.

For suppose that $a \approx b$. Given Beta-Eta Equivalence, this implies that $\forall X(Xa \leftrightarrow Xb)$, and given Universal Instantiation, this in turn implies that $(\lambda x.\varphi[x/a])a \leftrightarrow (\lambda x.\varphi[x/a])b)$. Finally, given Beta-Eta Equivalence, it follows that $\varphi \leftrightarrow \varphi[b/a]$.

In light of this fact, one might have the following worry. Given that Leibniz equivalence satisfies the orthodox postulates governing identity, isn't it a better candidate to be the identity relation than the relation that the Classical Opacitist takes to be identity? In a slogan, isn't Leibniz equivalence "more identity-like than identity"?[42]

The Classical Opacitist has a natural reply to this objection. Since Hesperus is identical to Phosphorus, any relation that does not relate Hesperus to Phosphorus is not identity-like. But given Universal Instantiation, Leibniz equivalence does not relate Hesperus and Phosphorus. For there are properties (for example, being known by the ancients to be visible at night) which Hesperus has but Phosphorus does not. In the crucial respect of its extension, Leibniz equivalence is simply not identity-like.

Let us now turn to a second style of objection to Classical Opacity. To fix ideas, suppose one accepted classical extensional mereology, and thought it was at least possible, and compatible with opacity of the Hesperus/Phosphorus variety, that the material world be comprised of a finite number of atomic particles (and fusions thereof), each of them discernible from all of the others by its spatial relations to the others. In such a world, it seems that every plurality of particles could be singled out in the language of physics, and hence every material object (or plurality thereof) could be singled out in the language

[42] See Bacon & Russell (2017, p. 4) for this objection and this quotation.

of physics plus mereology. Suppose, moreover, that all predicates of this language are transparent. Then it seems that Classical Opacity must be false. Let P be the property of being expressible in the language of physics plus mereology. If all predicates of this language are transparent, then Classical Opacitists should think that P properties never distinguish identical individuals: $\forall X \forall x \forall y(PX \wedge x = y \rightarrow (Xx \leftrightarrow Xy))$. Let M be the property of being a material object. The idea that we can single out all such objects in the language of physics plus mereology is then naturally regimented as the claim: $\forall X \exists Y(PY \wedge \forall x(Mx \rightarrow (Xx \leftrightarrow Yx)))$. But given classical quantification theory (and the assumption that anything identical to a material object is a material object) it follows that: $a = b \wedge Ma \rightarrow (Fa \leftrightarrow Fb)$. Since Hesperus is a material object, this rules out Hesperus/Phosphorus-style opacity.

We think that Classical Opacitists should not be worried by this argument. In particular, they should reject the above articulation of the idea that, for any condition on material objects, there is a physical-mereological (and hence transparent) property coextensive with it, in the sense of applying to all and only the things that have it. What they should say instead is that, in such a world, for any condition on material objects, there is a physical-mereological property *weakly coextensive* with it, in the sense of applying to all and only the things that are *identical to something* that has it: $\forall X \exists Y(PY \wedge \forall x(Mx \rightarrow (\exists\, y(y = x \wedge Xy) \leftrightarrow Yx)))$.

There are two more general morals here. First, in most scientific applications we are used to working with transparent predicates and appealing freely to Substitution. In such applications, claims of coextensiveness and claims of weak coextensiveness come to the same thing. But according to Classical Opacitists, such claims do not come to the same thing once we start reckoning with opacity. In general, it is weak coextensiveness, not coextensiveness, that Classical Opacitists should use in theorizing about entities that are, intuitively, 'individuated extensionally', such as material objects (if classical extensional mereology is true), sets, pluralities, and so on. Second, denying the coextensiveness of identity and Leibniz equivalence doesn't doom the project of characterizing identity in independent terms. For example, identity between individuals is arguably coextensive with being members of the same sets/belonging to the same pluralities.

Let's say that a *non-classical* theory of opacity is a theory that is consistent with opacity but which rules out opacity given Universal Instantiation, and so cannot be extended to a theory of Classical Opacity. A third, quite different style of objection to Classical Opacity maintains that certain non-classical theories of opacity are to be preferred to theories of classical opacity on the basis of general theoretical virtues such as strength, simplicity and parsimony. To illustrate this style of objection we'll now develop one natural family of theories motivated by the desire to preserve Leibniz's Law. Such theories give formal expression to the slogan that opaque predicates "don't express genuine properties". We think these theories are independently interesting, but our main goal will be to compare them to theories of Classical Opacity, and to argue that theories of Classical Opacity are on balance more attractive. Our discussion will not be exhaustive – we can't canvass the full space of theories consistent with opacity – but the considerations adduced below are general enough that we expect them to apply to many alternative theories we might have considered instead.

If there are cases of opacity, then, given the Core Theory, some sentence of the form $a = b \rightarrow (Fa \leftrightarrow Fb)$ is false. If such a sentence is false, and the corresponding instance of Leibniz's Law ($a = b \rightarrow \forall X(Xa \leftrightarrow Xb)$) is true, then the corresponding instance of Universal Instantiation must be false. In the case of first-order languages, perhaps the most

familiar way of motivating failures of Universal Instantiation comes from names like "Zeus". Positive free logicians claim that although Zeus is a mythical god, there are no mythical gods, and indeed there is no such thing as Zeus. An opacitist who wishes to preserve Leibniz's Law might think of opaque predicates by analogy to the way in which positive free logicians think of names like "Zeus". They might thus maintain that although Hesperus was known by the Greeks to be visible at night and Phosphorus was not known by the Greeks to be visible at night, there is nothing that Hesperus is that Phosphorus is not. To handle cases names like "Zeus" in first-order languages, positive free logicians reject Universal Instantiation and replace it with a principle restricted by an existence proviso:

Free Instantiation $\exists x(x = a) \rightarrow (\forall x\varphi \rightarrow \varphi[a//x])$.[43]

In the remainder of this section, we'll explore theories which do not include all instances of Universal Instantiation but do include every instance of Free Instantiation.

Adopting Free Instantiation in place of Universal Instantiation does not on its own guarantee Leibniz's Law. To ensure that Leibniz's Law is a theorem we impose the following principle:

Free Substitution $\exists X(X = F) \rightarrow (a = b \rightarrow Fa = Fb)$.

Free Substitution says that existent entities are transparent. Together with Free Instantiation it implies Leibniz's Law against the background of the Core Theory.[44]

These principles alone constitute an unsatisfyingly weak theory of identity. For example, they leave open the bizarre possibility that applying an existent entity to another existent entity could yield a non-existent entity. This could happen for example if applying a transparent existent entity to a transparent existent entity yielded an opaque, and hence non-existent, entity. It is not just bizarre that existence could fail to be closed under application. Together, these principles also leave open that the following principle could fail:

Generalized Leibniz's Law $a = b \rightarrow \forall X\forall x((Xx)a \leftrightarrow (Xx)b)$.

But those attracted to Leibniz's Law will naturally find Generalized Leibniz's Law attractive too.

We will therefore strengthen the theory by imposing in addition the following principle:

Existence Preservation $\forall X\forall x\exists Z(Z = Xx)$.

Existence Preservation ensures that existence is closed under application. Free Instantiation, Free Substitution and Existence Preservation together also entail Generalized Leibniz's Law. In this argument, Existence Preservation plays a role similar to that of Transparency Preservation in Section 3.1. Since all existent entities are transparent (by

---

[43] See, for example, Bacon (2013).

[44] *Proof:* By Free Substitution and Generalization, we have: $\forall X(\exists Z(Z = X) \rightarrow a = b \rightarrow Xa \leftrightarrow Xb)$. Distributing the quantifiers over the conditional, we get: $\forall X\exists Z(Z = X) \rightarrow \forall X(a = b \rightarrow Xa \leftrightarrow Xb)$. But we have as a theorem: $\forall X\exists Z(Z = X)$, and so it follows that we have: $\forall X(a = b \rightarrow Xa \leftrightarrow Xb)$. Leibniz's Law follows by distributing the quantifier and eliminating its vacuous occurrence.

Free Substitution), Existence Preservation entails that applying an existent property or relation to an existent entity yields a transparent entity.

Since there is something identical to Hesperus, Free Instantiation allows us to instantiate any first-order generalization with "Hesperus" (and likewise for "Phosphorus"). So if Hesperus/Phosphorus is a case of opacity, then the corresponding instance of Quantified Substitution must be false. More generally, the adoption of Free Substitution (and hence the retreat from Universal Instantiation to Free Instantiation) was motivated by the idea that opaque predicates are empty in the way that positive free logicians traditionally take names like "Zeus" to be empty. Nothing was said to impugn Universal Instantiation for expressions of type e or type t, which cannot be opaque because they do not take arguments. Thus it is natural to impose the principle:

e/t-**Existence** $\exists x(x = \alpha)$ if α is of type e or t.

In imposing this principle, we are confining our attention to misgivings about Universal Instantiation due to considerations related to opacity, and setting aside misgivings due to considerations related to empty names and the like. Together with Free Instantiation, this ensures that we will have Universal Instantiation for expressions of type e and type t.

Theories of *Free Opacity* are theories that are consistent with opacity and that include the Core Theory, Free Instantiation, Free Substitution, Existence Preservation and e/t-Existence.

Free Opacititists, like Classical Opacitists, must choose between Lift Congruence and Application Congruence, since they accept the Core Theory in which those principles together entail Substitution. As in the case of Classical Opacity, both choices are possible.

Let *Free Applicativism* be the theory axiomatized by Application Congruence, Constant Transparency, Transparency Preservation, Free Instantiation, Free Substitution, Existence Preservation, e/t-Existence and the Core Theory. In Appendix D, we show that Free Applicativism is consistent with opacity.[45]

The simplest way of developing an analogue of Classical Purity would be to add Purity to the principles of Free Opacity. Given the weaker background logic of Free Opacity, however, we can add a powerful additional axiom schema to this package while still allowing for cases of opacity.[46] Say that a sentence is *orthodox* just in case it is pure and is a theorem of transparent higher-order logic. Now consider the principle:

---

[45] In fact we show something stronger. Given Free Substitution, axioms which state the existence of logical operators are stronger than axioms which assert their transparency. Constant Transparency and Transparency Preservation follow respectively from the following two principles:

Constant Existence $\exists X(X = O)$ if O is an extended logical constant.

Continued Existence $\exists X(X = O)a$ if O is an extended logical constant.

We show that these principles are consistent with opacity given Free Applicativism.

[46] We here take inspiration from the "Pure Truth" schema of Bacon & Russell (2017). Their notion of "Pure Truth" differs from our notion of "orthodox" sentence in three ways. First, they allow open formulae to be Pure Truths. Second, they define Pure Truths model-theoretically, as opposed to axiomatically. Third, they define the Pure Truths in terms of a class of models which validate Functionality; thus $\forall X \forall Y(\forall x(Xx = Yx) \rightarrow X = Y$ is a Pure Truth, but not an orthodox sentence. This last difference is not significant; we will in fact give a consistency result for a theory which includes this principle as a theorem as well.

Orthodoxy φ, provided that φ is orthodox.

Let *Free Purity* be the smallest theory that contains Orthodoxy, Purity, Propositional Conjunction Congruence, Identity Congruence, Free Instantiation, Free Substitution, e/t-Existence and the Core Theory.[47] In Appendix D, we show that Free Purity is consistent with opacity.[48] It is an open question whether Free Purity is consistent with Conjunction Congruence.

We earlier discussed a number of tradeoffs between Classical Applicativism and Classical Purity. In the case of theories of Free Opacity, however, we think the balance is firmly on the side of Free Purity. A first reason why is that Free Purity includes Orthodoxy. Orthodoxy cannot be added to Free Applicativism; together with Application Congruence, Free Instantiation, and Free Substitution, Orthodoxy implies that our motivating examples are not cases of opacity. The sentence $\forall x \forall y(x = y \rightarrow (\lambda Z.Zx) = (\lambda Z.Zy))$ is orthodox. Given Application Congruence, it implies that $\forall x \forall y(x = y \rightarrow (\lambda Z.Zx)F = (\lambda Z.Zy)F)$. Given Reduction Congruence and Material Equivalence, this implies that $\forall x \forall y(x = y \rightarrow Fx = Fy)$. But we've already seen that e/t-Existence together with this principle would rule out cases of opacity at type e and type t. In particular it would rule out Hesperus/Phosphorus, since Hesperus (i.e. Phosphorus) exists.

A second reason that the balance of considerations tells in favor of Free Purity is that this theory avoids our central limitative results for Classical Purity. Although Lemma 1 and Propositions 2 and 3 can still be derived using Free Purity in place of Classical Purity, deriving problematic consequences from those claims requires Universal Instantiation. For example, Free Purity and Booleanism entail that $p = q \rightarrow (p \rightarrow p) \approx (p \rightarrow q)$. But in the theory of Free Purity, we cannot derive from this the result that the Greeks knew that Hesperus is visible at night if Hesperus is visible at night only if they knew that Phosphorus is visible at night if Hesperus is visible at night. For we cannot instantiate on the higher-order universal generalization in the definition of $\approx$ without a side-premise stating that the property of being known by the Greeks exists, which proponents of Free Opacity will deny.

In Appendix D we show that Free Purity and Booleanism are consistent with opacity.

Free Purity therefore seems to us the more attractive theory of Free Opacity. But is it more attractive than our theories of Classical Opacity? More specifically, is upholding Orthodoxy worth the cost of relinquishing Universal Instantiation?[49]

Bacon & Russell (2017) suggest that Orthodoxy is motivated as a form of conservativism. They argue that even if there is opacity, it should be undetectable in the language of pure logic. But we see no good reason to adopt this conservative methodology. It runs counter to the methodology of natural science, where we are open to new discoveries having revisionary ramifications for established theories. For example, physicists search for new kinds of particles precisely because they hope such discoveries will suggest ways

---

47 Existence Preservation follows from Orthodoxy, since all of its instances are orthodox.

48 Instead of Purity, we could consider the principle

Pure Existence $\exists x(x = a)$, provided a is pure.

In the presence of Free Substitution this implies Purity, but is stronger than it. The Appendix also shows that this principle is consistent with opacity given Free Purity.

49 The two schemas jointly imply Substitution against the background of the Core Theory, since the universal closure of any instance of Substitution is orthodox.

of modifying our existing theories of the particles we already know about. The situation seems similar to us in the case of logic. If we suppose that there is opacity, then there is a reason to question at least certain "laws" of pure logic – in particular the putative law $\forall x \forall y(x = y \rightarrow \forall X(Xx \leftrightarrow Xy))$.

An abductive argument for Orthodoxy is more promising. Perhaps the gain in strength it affords theories of Free Purity outweighs the loss of strength concomitant with giving up Universal Instantiation. This suggestion might be sharpened as follows. Let the first-order fragment of a higher-order language be those formulae of the language whose variables are all of type e and whose constants, other than the logical constants $\forall_e$ and $=_e$, $\wedge^t$ and $\neg^t$ are of type e, $e \rightarrow t$ and $e \rightarrow e \rightarrow \ldots \rightarrow e \rightarrow t$. Since we have assumed e/t-Existence as a component of Free Opacity, the theorems of the weakest theory of Classical Opacity contained in the first-order fragment of our language are just the theorems of the weakest theory of Free Opacity contained in that fragment. Someone might argue that the loss of 'higher-order' Universal Instantiation is a small cost. Perhaps they feel that while instantiation at types e and t is sacrosanct, instantiation at higher types is less compelling. And, in particular, they might think the cost of giving it up is outweighed by the gain in strength afforded by Orthodoxy.

We think this thought is mistaken, and betrays insufficient attention to the importance of quantificational reasoning in a range of settings beyond first-order theorizing. A striking example arises in connection to the version of functionalism according to which propositional attitudes like belief, desire, etc. are the unique relations between individuals and propositions that jointly satisfy a certain theoretical role. This view depends on being able to existentially generalize into predicate position; i.e., to instantiate variables with the very psychological predicates that are the most familiar candidates for opaque expressions. These are precisely the instances of Universal Instantiation that proponents of Free Opacity are committed to rejecting.[50]

Indeed, Free Opacity goes along with the view that there are simply no such relations as belief, hope, desire, etc. It is a kind of eliminativism about the propositional attitudes. Rather than rehearse standard arguments against such views, let us simply register our judgment that this constitutes a cost of Free Opacity far greater than any benefit afforded by Orthodoxy that we have been able to identify.

5 Quantified Substitution

In this section, we continue our defense of Classical Opacity by considering and rebutting two arguments in favor of Quantified Substitution, and therefore against Classical Opacity (and Free Opacity).

There is a tradition of claiming that Quantified Substitution is analytic. Quine for example writes "My position is that we can settle objectively and absolutely what predicate of a theory to count as the identity predicate, if any, once we have settled what notations to count as quantifiers, variables, and the truth functions" (Quine, 1961, p. 325). He explains that once we have settled these facts about notation, we can merely check whether a predicate

[50] In this connection Lewis (1970, p. 429) writes: "We must assume that all occurrences of T-terms in the postulate of T are purely referential, open to existential generalization and to substitution by Leibniz's law." The point about existential generalization is the crucial one; the point about Leibniz's law can be rejected by the Classical Opacitist functionalist by using Leibniz equivalence rather than identity to characterize the sense in which the realization of a theory is unique.

satisfies Quantified Substitution and reflexivity, arguing that "These requirements fix identity uniquely" (326). Kaplan (1986, p. 275) claims that Quantified Substitution stands or falls with $\forall x(\exists y(y = x \wedge \varphi) \rightarrow \forall y(y = x \rightarrow \varphi))$, and writes that this principle can be "usefully applied backwards to test whether an identity sign signs identity."

Although these pronouncements are offered without argument, one can extract the following line of reasoning in favor of Quantified Substitution from Quine and Kaplan's overall discussion. First, it is claimed that the semantic value of a (first-order) variable on an assignment is an object. Second, it is claimed that given this "objectual" semantics for variables, Quantified Substitution must hold (see e.g. (Kaplan, 1986, p. 244)).[51]

This line of reasoning seems to be running together model theory and semantics in precisely the way we cautioned against in Section 3.1. In particular, it seems to be invoking the model-theoretic notion of the value of a variable relative to an assignment function to draw conclusions about the meaning of universal generalizations. There are a number of ways to see that this line of reasoning is unsound. One is to observe that we ourselves appeal to assignment functions in the appendices of this paper to give models of Classical Opacity. Another is to note that higher-order theories of the sort we are describing are equivalent to "variable-free" theories in which certain combinators are taken as logical constants; variable-theoretic considerations simply do not arise when these theories are reformulated in this way. The argument fares little better if Quine and Kaplan's talk of variables having values is intended to be a robustly semantic notion, for it seems that only a free variable could sensibly be said to have an individual as its *meaning*, in which case the implications for generalizations (in which the relevant variables are bound) are far from clear. This gap in the argument is amplified by the observation that, if "... is the semantic value of ... (relative to assignment ...)" is a genuinely semantic notion, rather than a bit of set theoretic model theory, then it may well be opaque: that Hesperus is the value of a variable need not guarantee that Phosphorus is the value of that variable. The denial of this claim is perfectly compatible with the dictum that the values of variables are individuals, since planets are individuals.[52]

The second argument for Quantified Substitution we will consider starts from an example. Recall that Hesperus is the second planet from the sun, and Phosphorus is too. Since Hesperus was known by the ancients to be visible at night, it follows, given either Classical or Free Opacity, that there is a planet second from the sun that was known by the ancients to be visible at night. And since Phosphorus was not known by the ancients to be visible at night, it follows, given either Classical or Free Opacity, that there is a planet second from the sun that was not known by the ancients to be visible at night. However, despite the fact that there is a planet second from the sun that was known by the ancients to be visible at night, and the fact that there is a planet second from the sun that was not known by the ancients to be visible at night, there is only one planet second from the sun.

This example shows that, given either Classical or Free Opacity, there will be counterexamples to the following pretheoretically attractive principles, presented below with schematic arguments in English, to which they correspond:

---

[51] For helpful discussion of these ideas, and a response to this argument based on considerations orthogonal to our main concerns here, see Richard (1987).

[52] Bacon & Russell (2017, Section 2) make a similar point about the semantics of proper names in a somewhat different context.

6. $(\exists x(Fx \wedge Gx) \wedge \exists x(Fx \wedge \neg Gx)) \rightarrow \exists x \exists y(x \neq y \wedge Fx \wedge Fy)$

(a) Something is both F and G.

(b) Something is both F and not-G.

(c) $\therefore$ There are at least two things that are F.

7. $(\exists x(Fx \wedge \forall y(Fy \rightarrow x = y)) \wedge \exists x(Fx \wedge Gx)) \rightarrow \forall x(Fx \rightarrow Gx)$

(a) There's exactly one thing that is F.

(b) Something is both F and G.

(c) $\therefore$ Everything that is F is G.

Now 6 and 7 follow given Quantified Substitution and extremely weak principles governing the quantifiers. Do these principles, then, provide good reason to favor a theory of identity consistent with opacity that includes Quantified Substitution? We think that the answer is no.

To see why first consider the argument that led us to consider 6 and 7. That argument depended on drawing the conclusion "there is a planet second from the sun that was known by the ancients to be visible at night" from the premise "Hesperus is a planet second from the sun and Hesperus was known by the ancients to be visible at night", and likewise the conclusion "there is a planet second from the sun that was not known by the ancients to be visible at night" from the premise "Phosphorus is a planet second from the sun and Phosphorus was not known by the ancients to be visible at night". Such inferences – which the proponent of Classical Opacity will accept as straightforwardly valid, and which the proponent of Free Opacity will accept as valid given the existence of Hesperus (Phosphorus) – are at least as compelling as the inferences corresponding to the principles 6 and 7. It follows that if there are cases of opacity, then either certain plausible quantificational inferences fail or the patterns corresponding to 6 and 7 fail. It is, however, far from clear that the right choice here is to uphold 6 and 7, and reject the quantificational inferences.

Indeed, given the plausibility of the quantificational inferences, we're inclined to think that, once one has countenanced the existence of opacity, 6 and 7 should seem considerably less plausible. For example, once it's been noted that, despite Hesperus and Phosphorus being one and the same planet, Hesperus was a planet second from the sun known by the ancients to be visible at night, while Phosphorus was not, one should not be inclined to conclude from there being only one planet second from the sun, and there being something, namely Hesperus, that was a planet second from the sun known by the ancients to be visible at night, that anything which is a planet second from the sun was known by the ancients to be visible at night. For Phosphorus is a planet second from the sun, and it was not known by the ancients to be visible at night.

Moreover, while rejecting 6 and 7 may seem initially unpalatable, we think that, on inspection, this option is significantly preferable to endorsing the sorts of claims that follow given Quantified Substitution and the existence of opacity. Recall that Free Instantiation is inconsistent with Quantified Substitution and Hesperus/Phosphorus (since Hesperus exists). Given our motivating cases of opacity, then, if one wants to endorse Quantified Substitution, one must

reject both Universal Instantiation and Free Instantiation. But rejecting Free Instantiation leads to some very strange claims. In particular, one must accept some claim of the form $\exists x(x = a) \wedge Fa \wedge \neg\exists xFx$. For example, one must accept either: (a) There is something identical to Hesperus, and Hesperus is a planet second from the sun known by the ancients to be visible at night, but nothing is a planet second from the sun known by the ancients to be visible at night, or: (b) There is something identical to Phosphorus, and Phosphorus is a planet second from the sun not known by the ancients to be visible at night, but nothing is a planet second from the sun not known by the ancients to be visible at night. Both of these claims strike us as bizarre. It is just very hard to understand how, for example, if Hesperus is something, and is a planet second from the sun which was known by the ancients to be visible at night, there could be nothing that is a planet second from the sun which was known by the ancients to be visible at night.

We conclude that while opacicists may consistently uphold 6 and 7 by endorsing Quantified Substitution, the costs of doing so outweigh the supposed benefits. The failures of 6 and 7 are not grounds for preferring a different theory of identity consistent with opacity over theories of Classical or Free Opacity. Rather, the failures of these principles simply highlight the way in which valid quantificational reasoning casts the radical nature of opacity into sharp relief.[53,54]

6 Conclusion

Traditionally, it has been held that even if there are cases of opacity, both Quantified Substitution and Leibniz's Law will still hold. In this paper, we have argued, to the contrary, that cases of opacity do not give us a decisive reason to reject Universal Instantiation, and that if there are cases of opacity, both Quantified Substitution and Leibniz's Law should be rejected.

Opacitists face a number of difficult decisions in developing their theory of identity. We began by identifying one crucial such choice, between Applicative Congruence and Lift Congruence. We then developed two theories of Classical Opacity, against the

---

[53] Another way to dramatize this point is that opacitists of every stripe must say that $Fa \wedge \forall x(Fx \rightarrow x = a)$ isn't equivalent to $\forall x(Fx \leftrightarrow x = a)$, despite both being equivalent given Substitution and neither being an unnatural regimentation of "a is the unique F".

[54] There is an additional challenge for views which uphold Quantified Substitution. They must say something systematic about which terms can be instantiated. Bacon & Russell (2017) suggest the following candidate

Pure Instantiation $\forall x\varphi \rightarrow \varphi[a//x]$, provided a is a pure term.

But they show that this principle is inconsistent with opacity, Quantified Substitution, and $(\lambda xyZ.\top) = (\lambda xyZ.x = y \rightarrow (Zx \leftrightarrow Zy))$, where $\top$ is any propositional tautology. An obvious generalization of their argument shows that if there are cases of opacity, then Pure Instantiation and Quantified Substitution imply:

Pure Universal Distinctness $(\lambda xyZ.(x = y) \rightarrow (Zx \leftrightarrow Zy)) \neq \alpha$ for every pure α such that $\alpha xyZ \in T$ for x, y, Z of appropriate types.

Pure Universal Distinctness, however, has false instances in the most familiar, consistent theories of the fineness of grain of propositions, properties and relations. So fans of Quantified Substitution must either give up Pure Instantiation (and so face the question of what instantiation principle they can accept) or adopt an unfamiliarly fine-grained conception of relations. We are skeptical that this way lies a comparatively strong and simple theory of identity.

background of our Core Theory, which responded to that choice in different ways. We thereby showed that one can develop strong, simple theories of Classical Opacity.

We then defended theories of Classical Opacity against a series of objections. In the service of those objections, we developed theories of Free Opacity, which hold on to Leibniz's Law, but reject Quantified Substitution. We argued that on balance these theories were less attractive than theories of Classical Opacity. We then considered some objections to Classical Opacity that motivated Quantified Substitution, and argued that these objections too were unsuccessful, and that endorsing Quantified Substitution came with substantial costs of its own.

Throughout the paper we have appealed to the naturalness of classical quantification theory, and taken it to be a virtue of Classical Opacity that it upholds it. We defended this stance in our discussion of functionalism and eliminativism toward the end of Section 4. Our own view is that there are further strong abductive considerations in favor of classical quantificational logic.[55] If opacitists must give up this theory that would be a serious cost of countenancing opacity. We hope to have shown that they are not forced to.

References

Bacon, Andrew. 2013. Quantificational Logic and Empty Names. *Philosophers' Imprint*, 13(24), 1–21.

——. 2018. The Broadest Necessity. *Journal of Philosophical Logic*, 47(5), 733–783.

——, & Russell, Jeffrey Sanford. 2017. The Logic of Opacity. *Philosophy and Phenomenological Research*.

Benzmüller, Christoph, Brown, Chad E., & Kohlhase, Michael. 2004. Higher-Order Semantics and Extensionality. *Journal of Symbolic Logic*, 69(4), 1027–1088.

Braun, David. 1998. Understanding belief reports. *The Philosophical Review*, 107(4), 555–595.

Burgess, John P. 2008. *Tarski's tort*. Cambridge University Press. Page 149–168.

Carnap, Rudolf. 1947. *Meaning and necessity: a study in semantics and modal logic*. University of Chicago Press.

Crimmins, Mark. 1992. *Talk about beliefs*. MIT Press.

——, & Perry, John. 1989. The prince and the phone booth: Reporting puzzling beliefs. *The Journal of Philosophy*, 86(12), 685–711.

Cumming, Sam. 2016 (May). *On an Alleged Ambiguity in Attitude Reports*. Unpublished MS.

Dorr, Cian. 2014a. Quantifier variance and the collapse theorems. *The Monist*, 97(4), 503–570.

——. 2014b. Transparency and the context-sensitivity of attitude reports. *Pages 25–66 of:* Garcia-Carpintero, Manuel, & Martí, Genoveva (eds), *Empty Representations: Reference and Non-existence*. Oxford University Press

——. 2016. To be F is to be G. *Philosophical Perspectives*, 30(1), 39–134.

Edgington, Dorthy. 2002. Williamson on Vagueness, Identity and Leibniz's Law. *Pages 305–318 of:* Giaretta, P., Bottani, A., & Marrara, M. (eds), *Individuals, Essence and Identity: Themes of Analytic Metaphysics*. Dordrecht, The Netherlands: Kluwer.

[55] See Williamson (2013), Dorr (2014a), Goodman (2016).

Evans, Gareth. 1978. Can There Be Vague Objects? *Analysis*, 38, 208.
Fine, Kit. 2003. The Non-Identity of a Material Thing and Its Matter. *Mind*, 112(446), 195–234.
Frege, Gottlob. 1892. Über Sinn und Bedeutung. *Zeitschrift für Philosophie und philosophische Kritik*, 100, 25–50.
Fritz, Peter, & Goodman, Jeremy. 2016. Higher-Order Contingentism, Part 1: Closure and Generation. *Journal of Philosophical Logic*, 45(6), 645–695.
Geach, Peter. 1967. Identity. *Review of Metaphysics*, 21, 2–12.
Gibbard, Allan. 1975. Contingent identity. *Journal of Philosophical Logic*, 4(2), 187–221.
Goodman, Jeremy. 2016. An Argument For Necessitism. *Philosophical Perspectives*, 30 (1), 160–182.
——. 2017. Reality is Not Structured. *Analysis*, 77(1), 43–53.
——. forthcoming. Agglomerative Algebras. *Journal of Philosophical Logic*.
——, & Lederman, Harvey. 2018. *Perspectivism*. Unpublished Manuscript.
——, & ——. 2019. Sense, Reference and Substitution. *Philosophical Studies*.
Gupta, Anil. 1980. *The Logic of Common Nouns: an investigation in quantified modal logic*. Yale University Press.
Hawthorne, John. 2003. Identity. *Pages 99–130 of:* Loux, Michael J, & Zimmerman, Dean W (eds), *The Oxford Handbook of Metaphysics*. Oxford Handbooks.
Heck, Richard G. 1998. That there might be vague objects (so far as concerns logic). *The Monist*, 81(2), 274–296.
Kaplan, David. 1968. Quantifying in. *Synthese*, 19, 178–214.
——. 1986. Opacity. *Pages 229–289 of:* Hahn, Lewis Edwin, & Schilpp, Paul Arthur (eds), *The Philosophy of W. v. Quine*. Open Court.
Kocurek, Alex. forthcoming. On the Substitution of Identicals in Counterfactual Reasoning. *Noûs*
Lewis, David. 1970. How to Define Theoretical Terms. *Journal of Philosophy*, 67(13), 427–446.
——. 1971. Counterparts of persons and their bodies. *The Journal of Philosophy*, 68(7), 203–211.
Magidor, Ofra. 2011. Arguments by Leibniz's law in metaphysics. *Philosophy Compass*, 6(3), 180–195.
Mates, Benson. 1952. Synonymity. *Pages 109–136 of:* Linsky, Leonard (ed), *Semantics and the Philosophy of Language*,. Urbana: University of Illinois Press.
Mitchell, John. 1996. *Foundations for Programming Languages*. MIT Press.
Priest, Graham. 2005. *Towards Non-Being*. Oxford University Press.
Prior, Arthur N. 1971. *Objects of Thought*. Oxford University Press.
Quine, Willard van Orman. 1943. Notes on existence and necessity. *The Journal of Philosophy*, 40(5), 113–127.
——. 1956. Quantifiers and propositional attitudes. *The Journal of Philosophy*, 53(5), 177–187.
——. 1961. Reply to Professor Marcus. *Synthese*, 13(4), 323–330.
——. 1960. *Word and Object*. MIT Press.
——. 1986. Reply to Sellars. *Pages 337–340 of:* Hahn, Lewis Edwin, & Schilpp, Paul Arthur (eds), *The Philosophy of W. v. Quine*. Open Court.
Rayo, Agustín. 2013. *The Construction of Logical Space*. Oxford: Oxford University Press.

Richard, Mark. 1987. Quantification and Leibniz's law. *The Philosophical Review*, 96(4), 555–578.
Ripley, David. 2012. Structures and Circumstances: Two Ways to Fine-Grain Propositions. *Synthese*, 189, 97–118.
Salmon, Nathan U. 1986. *Frege's puzzle*. MIT Press.
Saul, Jennifer M. 1997. Substitution and simple sentences. *Analysis*, 57(2), 102–108.
——. 2010. *Simple sentences, substitution, and intuitions*. OUP Oxford.
Schiffer, Stephen. 1979. Naming and knowing. *Pages 28–41 of:* Peter, A. French, Theodore, E. Uehling, Jr, Howard, & Wettstein, K. (eds), *Midwest Studies in Philosophy II: Contemporary Perspectives in the Philosophy of Language*. University of Minnesota Press.
Soames, Scott. 1987. Direct reference, propositional attitudes, and semantic content. *Philosophical Topics*, 15(1), 47–87.
Stalnaker, Robert. 2012. *Mere Possibilities*. Princeton University Press.
Williamson, Timothy. 2002. Vagueness, Identity and Leibniz's Law. *In:* Giaretta, P., Bottani, A., & Marrara, M. (eds), *Individuals, Essence and Identity: Themes of Analytic Metaphysics*. Dordrecht, The Netherlands: Kluwer.
——. 2003. Everything. *Philosophical Perspectives*, 17, 415–465.
——. 2013. *Modal Logic as Metaphysics*. Oxford University Press.

Appendix A: Model Theory for Higher Order Logic

We begin by reviewing some standard definitions and concepts. The notation introduced here will be used throughout the appendices.

Definition 4: The set of types $\mathcal{T}$ is the smallest set such that $e, t \in \mathcal{T}$ and such that if $\sigma, \tau \in \mathcal{T}$ and $\tau \neq e$, then $\sigma \rightarrow \tau \in \mathcal{T}$.

Definition 5: A *typed family of sets* is a function mapping types to sets.

A typed family of sets F is *disjoint* just in case $F(\sigma) \cap F(\tau) = \emptyset$ whenever $\sigma \neq \tau$.

A typed family of sets F is *universally rich* just in case for all σ, $F(\sigma)$ is countably infinite.

Given a typed family of sets F and a type σ, for readability we often write F_σ in place of $F(\sigma)$. In the special case where F_σ is a singleton for all types σ, we abuse notation slightly, writing F_σ for the single element of F_σ.

Definition 6: A function f is a *typed family of functions* from the typed family of sets D to the typed family of sets E just in case it is a function on types such that for any type $\sigma f(\sigma)$ is a function from D_σ to E_σ.

For a typed family of functions f we often write f_σ for readability in place of $f(\sigma)$. If the domain D of a typed family of functions f is disjoint we often speak as if f is a function with domain $\bigcup_{\sigma \in \mathcal{T}} D_\sigma$, writing simply $f(x)$ in place of $f_\sigma(x)$ where it is clear from context that $x \in D_\sigma$.

Definition 7: An *applicative structure* is an ordered pair $\langle D, @\rangle$, where D is a typed family of sets, and $@$ is a typed family of functions such that for any type $\sigma \to \tau$, $@_{\sigma\to\tau}$ is a function which maps elements of $D_{\sigma\to\tau}$ to elements of $D_\tau^{D_\sigma}$.

One can also think of $@$ as a (curried) function of two arguments, which first takes an element of $D_{\sigma\to\tau}$, then an element of D_σ and produces an element of D_τ. In line with this way of thinking, infix notation is sometimes used for $@$, with $f@x$ meaning $@(f)(x)$. Following this convention we typically write $f@$ instead of $@f$.

Definition 8: An applicative structure is *functional* just in case for every σ and τ, $@_{\sigma\to\tau}$ is injective, i.e. for all $x, y \in D_{\sigma\to\tau}$, if $x \neq y$, then $x@ \neq y@$.

An applicative structure is *full* just in case for every σ and τ, $@_{\sigma\to\tau}$ is surjective, i.e. for every $f \in D_\tau^{D_\sigma}$, there is some $x \in D_{\sigma\to\tau}$ such that $x@ = f$.

Any functional applicative structure $\langle D, @\rangle$ is isomorphic to an applicative structure such that $D_{\sigma\to\tau} \subseteq D_\tau^{D_\sigma}$ and $@_{\sigma\to\tau}$ is the identity function Id_D. In discussing such applicative structures, then, we can restrict our attention to ones with these properties.

Any full and functional applicative structure is isomorphic to an applicative structure such that $D_{\sigma\to\tau} = D_\tau^{D_\sigma}$ and $@_{\sigma\to\tau}$ is the identity function.

Definition 9: *A signature* is a typed family of sets Σ. The elements of Σ_σ are the *constants* of type σ.

Definition 10: Fix a disjoint, universally rich typed family of sets V (the *variables*). The *terms* T^Σ of the signature Σ with variables V is the smallest typed family of sets satisfying the following constraints:

- If $x \in V_\sigma$, then $x \in T_\sigma^\Sigma$;
- If $a \in \Sigma_\sigma$, then $a \in T_\sigma^\Sigma$;
- If $F \in T_{\sigma\to\tau}^\Sigma$ and $a \in T_\sigma^\Sigma$, then $Fa \in T_\tau^\Sigma$;
- If $x \in V_\sigma$, and $F \in T_\tau^\Sigma$, then $\lambda x.F \in T_{\sigma\to\tau}^\Sigma$.

Throughout we assume fixed a disjoint universally rich V for variables. We generally assume the signature is clear from context, and typically write simply T instead of T^Σ. Where the signature is clear from context, we indicate that $a \in T_\sigma$ by writing a^σ.

Definition 11: Given an applicative structure $A = \langle D, @\rangle$ an *assignment function* g is any typed family of functions from V to D. An *interpretation function* for A, $[\![\cdot]\!]$, is a function mapping assignment functions to typed families of functions from T to D that satisfy the following constraints:

- $[\![v]\!]^g = g(v)$, for each variable $v \in \bigcup_{\sigma\in\mathcal{T}} V_\sigma$;
- $[\![a^\sigma]\!]^g = [\![a^\sigma]\!]^{g'}$ and $[\![a^\sigma]\!]^g \in D_\sigma$ for all $\sigma \in \mathcal{T}$, all constants $a^\sigma \in \Sigma_\sigma$, and all variable assignments g, g';

- $[\![\alpha\beta]\!]^g = [\![\alpha]\!]^g @ [\![\beta]\!]^g$;

- $[\![\lambda x^\sigma.\alpha]\!]^g @ d = [\![\alpha]\!]^{g[x \mapsto d]}$, where in general $g[x \mapsto z]$ is the assignment function such that $g[x \mapsto z](x) = z$ and for all other $y \in \bigcup_{\sigma \in \mathcal{T}} V_\sigma$, $g[x \mapsto z](y) = g(y)$.

Definition 12: An *applicative model* is a structure $\langle D, @, [\![\cdot]\!], v\rangle$ where $A = \langle D, @\rangle$ is an applicative structure, $[\![\cdot]\!]$ is an interpretation function for A, and $v : D_t \to \{0, 1\}$ maps elements of D_t to truth-values.

Definition 13: A formula φ is *valid* on an applicative model $\langle D, @, [\![\cdot]\!], v\rangle$ if and only if $v([\![\varphi]\!]^g) = 1$ for all assignment functions g. A theory Γ is valid on a model just in case every $\varphi \in \Gamma$ is valid on the model.

Definition 14: A signature Σ is *logical* just in case $\neg \in \Sigma_{t \to t}$, $\wedge \in \Sigma_{t \to t \to t}$, and for all σ, $\forall_\sigma \in \Sigma_{(\sigma \to t) \to t}$.

A signature is a signature *with primitive identity* just in case $=_\sigma \in \Sigma_{\sigma \to \sigma \to t}$, for each type σ.

In proving the consistency of theories we are interested in, we will always use logical signatures with primitive identity.

Definition 15: An applicative model $\langle D, @, [\![\cdot]\!], v\rangle$ for a logical signature is *standard* if and only if for all assignment functions g:

- $v([\![\varphi]\!]^g) = 1$ iff $v([\![\neg \varphi]\!]^g) = 0$;

- $v([\![\varphi \wedge \psi]\!]^g) = 1$ iff $v([\![\varphi]\!]^g) = 1$ and $v([\![\psi]\!]^g) = 1$;

- $v([\![\forall_\sigma F]\!]^g) = 1$ iff $v([\![F]\!]^g @ a) = 1$ for all $a \in D_\sigma$.

It is a well-known fact, and one we will take for granted in what follows, that classical higher-order logic is valid on all standard models.

Definition 16: A model $\langle D, @, [\![\cdot]\!], v\rangle$ for a logical signature with primitive identity is *normal* just in case it is standard and $([\![=_\sigma]\!] @ x @ y) = 1$ if and only if $([\![=_\tau]\!] @ (f @ x) @ (f @ y)) = 1$ for every τ and every $f \in D_{\sigma \to \tau}$.

Transparent higher-order logic is valid on normal models.

From now on, we will simply say "applicative model" when we mean "applicative model for a logical signature with primitive identity."

A very simple class of applicative structures will be of interest in what follows, in part because they give rise to a simple, tractable class of standard models:

Definition 17: An applicative structure $\langle D, Id_D\rangle$ is an *intensional applicative structure* just in case for some pair $\langle E, W\rangle$ where E is a set, thought of as the set of individuals, and W is a non-empty set:

- $D_t = \mathcal{P}(W)$;

- $D_e = E$;
- $D_{\sigma\to\tau} = D_\tau^{D_\sigma}$ for all types σ and τ.

We say that the intensional applicative structure $\langle D, Id_D\rangle$ is *based on* such a pair $\langle E, W\rangle$.

This definition imposes more stringent conditions on an intensional applicative structure than one sometimes finds. First, we require that D_t be identical to $\mathcal{P}(W)$ rather than merely being a subset of it. Second, we require that higher domains $D_{\sigma\to\tau}$ be identical to $D_\tau^{D_\sigma}$, rather than merely being subsets of it. We have used this more restrictive definition because our constructions below only employ intensional applicative structures as defined here.

Any intensional applicative structure is based on a unique pair $\langle E, W\rangle$. In what follows we will often use "intensional applicative structure" directly to describe an appropriate pair $\langle E, W\rangle$. Strictly speaking, what is meant is the unique intensional applicative structure $\langle D, Id_D\rangle$ which is based on $\langle E, W\rangle$.

Definition 18: An applicative model $\langle E, W, [\![\cdot]\!], v\rangle$ is an *intensional model* just in case $\langle E, W\rangle$ is an intensional applicative structure and

- $[\![\neg]\!]^g = \lambda x : x \in D_t . W - x$;
- $[\![\wedge]\!]^g = \lambda xy : x, y \in D_t . x \cap y$;
- $[\![\forall_\sigma]\!]^g = \lambda z : z \in D_{\sigma\to t} . \bigcap_{x\in D_\sigma} zx$;
- for some $w \in W$, for every $p \in D_t$, $v(p) = 1$ just in case $w \in p$.

As defined, all intensional models are functional and full. It is not built into the definition of an intensional model how identity will be interpreted. In later sections we will provide various ways of interpreting identity in these models.

The following schema is valid on intensional models:

$\approx$ **Booleanism** $a \approx b$, whenever $a \leftrightarrow b$ is a theorem of propositional logic.

This will be of interest in our later discussion.

Appendix B: Models of Classical Applicativism

We now turn to the key definitions for giving models of Classical Applicativism.

Definition 19: Fix an applicative model $M = \langle D, @, [\![\cdot]\!], v\rangle$. A typed family of sets $\sim$ is an *applicative congruence* for M just in case for all σ, $\sim_\sigma \in D_{\sigma\to\sigma\to t}$ and for every σ, τ and g (writing $x \sim_\sigma y$ for readability in place of $(\sim_\sigma @x)@y$):

- The relation which holds between x, y just in case $v(x \sim_\sigma y) = 1$ is an equivalence relation;
- $v(f \sim_{\sigma\to\tau} g) = 1$ just in case $v(f@x \sim_\tau g@x) = 1$ for every $x \in D_\sigma$;

- If $v(x \sim_t y) = 1$, then $v(x) = v(y)$;
- If $v(x \sim_t y) = 1$, then $v([\![\neg]\!]^g @x \sim_t [\![\neg]\!]^g @y) = 1$;
- If $v(x_1 \sim_t y_1) = 1$ and $v(x_2 \sim_t y_2) = 1$ then $v(([\![\wedge]\!]^g @x_1) @x_2 \sim_t ([\![\wedge]\!]^g @y_1) @y_2) = 1$;
- If $v(f \sim_{\sigma \to t} g) = 1$ then $v([\![\forall_\sigma]\!]^g @f \sim_t [\![\forall_\sigma]\!]^g @g) = 1$;
- $v(x_1 \sim_\sigma y_1) = 1$ and $v(x_2 \sim_\sigma y_2) = 1$ then $(x_1 \sim_\sigma x_2) \sim_t (y_1 \sim_\sigma y_2)$.

The basic idea is to use an applicative congruence to interpret identity.

Definition 20: A structure $\langle D, @, [\![\cdot]\!], v, \sim \rangle$ is a *congruential applicative model* if and only if $M = \langle D, @, [\![\cdot]\!], v \rangle$ is a standard applicative model, $\sim$ is an applicative congruence for M and for all σ and g, $[\![=_\sigma]\!]^g = \sim^\sigma$.

Proposition 21: *In addition to classical higher-order logic, Functionality, Application Congruence, Constant Transparency and Transparency Preservation are valid on every congruential applicative model.*

Proof: The validity of Functionality and Application Congruence follows straightforwardly from the fact that $\sim$ is an intensional applicative congruence. For Constant Transparency and Transparency Preservation, we must show that $\sim$ is a congruence with respect to the extended logical operators at all relational types. For given this, for each extended logical constant, one can show that it is transparent, and then (if applicable) that it satisfies Transparency Preservation. In the case of $\neg$ and $\wedge$ at higher relational types, we show that $\sim$ is a congruence with respect to them by an easy induction. The other cases are built into the definition. □

Thus, if there are any congruential applicative models of this kind in which not all instances of Substitution are valid, then Classical Applicativism is consistent with opacity.

The simplest way of constructing such models is to consider intensional models with equivalence relations E_e and E_t on types e and t respectively. If E_e or E_t holds between some x, y, then we let $x \sim_\sigma y = W$; otherwise $x \sim_\sigma y = \emptyset$, where $\sigma = e$ or $\sigma = t$ (this guarantees the last condition on an applicative congruence). We extend the $\sim_\sigma$ relations to higher types by letting $x \sim_{\sigma \to \tau} y = W$ if $x(z) \sim_\tau y(z)$ for all $z \in D_\sigma$ and $x \sim_{\sigma \to \tau} y = \emptyset$ otherwise. To ensure that this is an applicative congruence, we must check that E_t obeys the constraint concerning the valuation of elements of D_t and the constraints concerning $\wedge$, $\neg$ and $\forall_\sigma$. Say that a model exhibits opacity at type σ iff $\forall_\sigma x \forall_\sigma y \forall F(x = y \to (Fx \leftrightarrow Fy))$ is not valid in the model. It is straightforward to check that any full intensional model in which $|D_e| > 2$, and $x \sim_e y$ for some $x \neq y$ exhibits opacity at type e. It is trivial to construct such models, since there are no constraints on $\sim_e$, and we can let E_t be model-theoretic identity, in which case $\sim_t$ trivially satisfies the needed conditions. But we can also produce models which exhibit opacity at all types including t. For if in addition to E_e being nontrivial, pE_tq for $p \neq q$, then the

model will in fact exhibit opacity at every type. It is easy to see that if E_t is a non-universal equivalence relation that relates only propositions which receive the same value under v, and is a congruence with respect to set-theoretic complement and arbitrary set-theoretic intersection on the (complete atomic) boolean algebra D_t in our intensional model, then the conditions for an applicative congruence will be satisfied by $\sim_t$ and hence satisfied at every type. Since there are nontrivial such congruences, there are models in the class which exhibit opacity at every type.

It is a standard fact that the validities on models such as these are closed under Generalization. Thus since the models exhibit opacity at every type, they fail to validate Substitution at every type. Since Classical Applicativism is valid on the models, this suffices to show that Classical Applicativism is consistent with opacity.

As noted at the end of the previous section, intensional models validate $\approx$ Booleanism. But it is easy to construct models of Classical Applicativism which exhibit opacity and do not validate such a coarse-grained theory of propositions under Leibniz equivalence. In any standard applicative model, define $\sim_t$ as mapping p, q to some fixed true proposition if and only if $v(p) = v(q)$. $\sim_e$ can be arbitrary so long as it determines an equivalence relation. There are then many ways of extending the relations to higher types that are consistent with the above constraints; the main work is done by having the right kind of relation at type t. This equivalence relation can thus be used to give a coarse-grained theory of propositional identity consistent with very fine-grained propositions under Leibniz equivalence. Most proponents of fine-grained propositions under $\approx$ will want to have more equivalence-classes of D_t under $=$ than we have in this model (where there are only two). These examples show that the only in principle barriers to giving such models are general limitative results about fine-grained propositions (mentioned in fn. 33), and not a product of special features of Classical Applicativism.

Appendix C: Models of Classical Purity and Classical Pristineness

In this section, we provide models of Classical Pristineness and Classical Purity. We first introduce some general concepts and techniques that we'll appeal to repeatedly in constructing such models, and provide a sketch of how our models will validate the core principles of Classical Pristineness and Classical Purity.

Definition 22: Fix an applicative structure $A = \langle D, @ \rangle$, and suppose $\sim$ is a typed family of sets such that for all σ, $\sim_\sigma \subseteq D_\sigma \times D_\sigma$. $\sim$ is a *logical relation* for A just in case, for each type $\sigma \rightarrow \tau$, $f \sim_{\sigma\rightarrow\tau} g$ just in case, for every $x, y \in D_\sigma$, if $x \sim_\sigma y$, then $f@x \sim_\tau g@y$.

A logical relation $\sim$ may fail to be an equivalence relation at certain types, even if it is an equivalence relation at base types. To interpret identity, we will generate an equivalence relation by appealing to certain *sets* of logical relations.

Definition 23: An *equivalence generator* for an applicative structure $A = \langle D, @ \rangle$ is an indexed set of logical relations for A, $\{\sim^i\}_{i\in I}$ such that for each type σ, $\bigcup_{i\in I} \sim^i_\sigma$ is an equivalence relation.

Definition 24: An applicative model $M = \langle D, @, [\![\cdot]\!], v \rangle$ *interprets identity by the equivalence generator* $\{\sim^i\}_{i\in I}$ just in case $\{\sim^i\}_{i\in I}$ is an equivalence generator for the

applicative structure $\langle D, @\rangle$ and for every assignment function $v([\![a = b]\!]^g) = 1$ iff $[\![a]\!]^g \sim_i [\![b]\!]^g$ for some $i \in I$.

Models which interpret identity by an equivalence generator have nice properties. One of these is as follows:

Proposition 25: *Fix an applicative model $M = \langle D, @, [\![\cdot]\!], v\rangle$ which interprets identity by the equivalence generator $E = \{\sim^i\}_{i\in I}$. Suppose that for some assignment function g, for all $i \in I [\![F]\!]^g \sim_i [\![F]\!]^g$. Then $v([\![a = b]\!]^g) = 1$ only if $v([\![Fa = Fb]\!]^g) = 1$.*

Proof: Since M interprets identity by the equivalence generator E, if $v([\![a = b]\!]^g) = 1$ then $[\![a]\!]^g \sim_j [\![b]\!]^g$, for some $j \in I$. But, since $[\![F]\!]^g \sim_i [\![F]\!]^g$, for every $i \in I$, in particular, $[\![F]\!]^g \sim_j [\![F]\!]^g$. And so, given the definition of a logical relation, $[\![F]\!]^g @ [\![a]\!]^g \sim_j [\![F]\!]^g @ [\![b]\!]^g$, and so $[\![Fa]\!]^g \sim_j [\![Fb]\!]^g$, from which it follows that $v([\![Fa = Fb]\!]^g) = 1$, as desired. □

Another important fact about logical relations that will play a central role in generating models of Classical Pristineness and Purity is the following:

Proposition 26: *Fix an applicative model $M = \langle D, @, [\![\cdot]\!], v\rangle$. If $\sim$ is a logical relation for $\langle D, @\rangle$, then for any assignment function g, if F^τ is a closed term and all constants c^σ which occur in F are such that $[\![c^\sigma]\!]^g \sim_\sigma [\![c^\sigma]\!]^g$, then $[\![F]\!]^g \sim_\tau [\![F]\!]^g$.*

Proof: This follows from Lemma 8.2.5 (the Basic Lemma) in Mitchell (1996). For a sketch of this proof see Exercise 8.2.12 in Mitchell (1996). □

Together Propositions 25 and 26 immediately give us the following corollary, to which we will appeal repeatedly in what follows:

***Corollary 27**: Fix an applicative model $M = \langle D, @, [\![\cdot]\!], v\rangle$, which interprets identity by the equivalence generator $\{\sim^i\}_{i\in I}$. For any assignment function g, if F is a closed term and all constants c^σ which occur in F are such that $[\![c^\sigma]\!]^g \sim^i_\sigma [\![c^\sigma]\!]^g$ for all $i \in I$, then if $v([\![a = b]\!]^g) = 1$, so too $v([\![Fa = Fb]\!]^g) = 1$.*

This corollary implies that for any closed combinator F and any terms a and b, $a = b \rightarrow Fa = Fb$ will be valid on any model that interprets identity by an equivalence generator. In what follows our strategy will be to construct equivalence generators which relate the interpretations of some or all of the logical constants to themselves. We will then exploit this corollary to show that pristine and pure terms are transparent.

C.1 Models of Classical Pristineness

Any model of Classical Purity is also a model of Classical Pristineness. Our later models of Classical Purity would therefore suffice to show that Classical Pristineness is consistent with opacity. However, we can give simple, flexible models of Classical Pristineness which are not models of Classical Purity. These models allow us to establish that Classical Pristineness and $\approx$ Booleanism are consistent with opacity at type t. (Later we will show that $\approx$ Booleanism and Classical Purity are inconsistent with opacity at type t.) The models will also be the basis for our proof that Free Purity is consistent with opacity.

Definition 28: Fix an applicative model $M = \langle D, @, [\![\cdot]\!], v\rangle$. An equivalence generator $\{\sim_i\}_{i\in I}$ for $\langle D, @\rangle$ is *pristine* for M just in case for every $i \in I$ and $\sigma \in \mathcal{T}$, $[\![\neg]\!] \sim_i [\![\neg]\!]$, $[\![\wedge]\!] \sim_i [\![\wedge]\!]$, and $[\![=_\sigma]\!] \sim_i [\![=_\sigma]\!]$.

An applicative model $M = \langle D, @, [\![\cdot]\!], v\rangle$ is pristine just in case it interprets identity by an equivalence generator $\langle D, @\rangle$ which is pristine for M.

It follows immediately from Corollary 27 that Pristineness is valid on any pristine model. The theory Classical Pristineness contains more principles than just Pristineness and classical higher-order logic. We will prove that the full theory is consistent with opacity again by considering a class of simple intensional models which validate those further principles.

Definition 29: A structure $\langle E, W, V, \sim\rangle$ is a *Kripke intensional structure* just in case

- $\langle E, W\rangle$ is an intensional applicative structure;
- $V \subseteq W$; and
- $\sim$ is a logical relation for $\langle E, W\rangle$ such that:
  - $\sim_e$ is an equivalence relation on E;
  - $p \sim_t q$ iff $p \cap V = q \cap V$.

Definition 30: A structure $\langle E, W, V, \sim, [\![\cdot]\!], v\rangle$ is a *Kripke pristine intensional model* just in case $\langle E, W, V, \sim\rangle$ is a Kripke intensional structure, $\langle E, W, [\![\cdot]\!], v\rangle$ is an intensional model, and $[\![=_\sigma]\!]^g = \lambda xz.\{w \in W : x \sim_\sigma z \text{ or } x = z\}$.

In Kripke pristine intensional models, the relation $\sim$ is an equivalence relation on the domains for base types. It is easily checked that it is therefore guaranteed to be a partial equivalence relation on the domains for higher types, that is, it will be symmetric and transitive, but may fail to be reflexive. Given this, if $\sim^0 = \sim$ and $\sim^1$ is model-theoretic identity, then $\{\sim^i\}_{i\in\{0,1\}}$ is an equivalence generator. Similarly, one can readily check that any relation $\sim$ satisfying the above conditions will be a pristine equivalence generator, and it is clear from the clause for identity that the model interprets identity by this equivalence generator. Moreover, Identity Congruence, Propositional Conjunction Congruence, Identifying Identities, and Booleanism are valid on any Kripke pristine intensional model. It is easy to construct such models which exhibit opacity at every type. So, Classical Pristineness is consistent with opacity.

Kripke pristine intensional models do not in general validate Purity, however:

Example 31: Consider a Kripke pristine intensional model $\langle W, E, V, \sim, [\![\cdot]\!], v\rangle$ where $W = V = \{w\}$, $E = \{a, b\}$, $\sim_e = E \times E$. Consider $f, g \in D_{e\to t}$ such that $f(a) = f(b) = W$ and $g(a) = g(b) = \emptyset$. Observe that $\sim_{e\to t} = \{\langle f, f\rangle, \langle g, g\rangle\}$, since all other $h \in D_{e\to t}$ do not satisfy the conditions to be related by a logical relation. Now consider $T, U \in D_{(e\to t)\to t}$ where $U(h) = W$ for all $h \in D_{e\to t}$, and $T(f) = T(g) = W$, while for all $h \in D_{e\to t} \setminus \{f, g\}$, $T(h) = \emptyset$. ("U" is for "universal"; T is for "transparent".) We have $T \sim_{(e\to t)\to t} U$, since $T(f) = U(f)$ and

$T(g) = U(g)$, and these are the only members of $D_{e\to t}$ related by $\sim_{e\to t}$ to any other member of $D_{e\to t}$. But it is not the case that $[\![\forall_{e\to t}]\!]U \sim_t [\![\forall_{e\to t}]\!]T$, since in the first case the value is W, while in the second the value is $\emptyset$. Given the clause for identity, this means that $\forall_{(e\to t)\to t} F \forall_{(e\to t)\to t} G(F = G \to (\forall x Fx \leftrightarrow \forall x Gx)$ will be false. Since the model satisfies Universal Instantiation, Purity will not be valid on the model.

Note that this same example, of T and U, shows that Application Congruence and Quantified Application Congruence are not valid on these models. Functionality, however, is valid on them.

We are unaware of additional constraints on these Kripke pristine intensional models that would ensure that they validate Purity while validating Universal Instantiation. We'll see later that if Universal Instantiation is weakened to Free Instantiation, then simple, elegant models which validate Purity can be given.

C.2 Models of Classical Purity

Definition 32: Fix an applicative model $M = \langle D, @, [\![\cdot]\!], v\rangle$. An equivalence generator $\{\sim_i\}_{i\in I}$ for $\langle D, @\rangle$ *is pure* for M if and only if it is pristine for M, and for each $i \in I$, and $\sigma \in \mathcal{T}$, $[\![\forall_\sigma]\!] \sim_i [\![\forall_\sigma]\!]$.

An applicative model $M = \langle D, @, [\![\cdot]\!], v\rangle$ is pure just in case it interprets identity by an equivalence generator for $\langle D, @\rangle$, which is pure for M.

It follows immediately from Corollary 27 that Purity is valid on any pure model. We will construct models of Classical Purity using one kind of pure equivalence generator, based on a group of pairs of permutations.

Given an applicative structure $\langle D, Id_D\rangle$, and a pair of permutations $\pi = \langle \pi_e, \pi_t\rangle$ of D_e and D_t respectively, we define the action of π on all $D_{\sigma\to\tau}$ as $\pi_{\sigma\to\tau}(f) := \pi_\tau \circ f \circ \pi_\sigma^{-1}$. Given this definition, the pair $\pi = \langle \pi_e, \pi_t\rangle$ induces a typed family of functions from D to D, which we will write π_σ. We will often speak directly of such a pair $\pi = \langle \pi_e, \pi_t\rangle$ as if it were a typed family of functions and write $\pi(x) = y$ for $\pi_\sigma(x) = y$ when $x \in D_\sigma$.

Given such a pair of permutations π, let $x \sim_\pi y$ mean that $\pi(x) = y$. It is easy to check that $x \sim_\pi y$ is a logical relation. But $\sim_\pi$ is not an equivalence relation unless the permutations on the domains associated with base types are each the identity.

To construct an equivalence generator from logical relations induced by pairs of permutations we use groups of pairs of permutations. Recall that G is a group of permutations of a set X just in case every $\pi \in G$ is a permutation of X, and if $\pi, \pi' \in G$, then $\pi^{-1} \in G$ and $\pi \circ \pi' \in G$. We can extend this idea to pairs of permutations straightforwardly:

Definition 33: A set G is *a group of pairs of permutations of X and Y* just in case for every $\langle \pi_1, \pi_2\rangle \in G$, π_1 is a permutation of X and π_2 is a permutation of Y, and if $\langle \pi_1, \pi_2\rangle, \langle \pi_1', \pi_2'\rangle \in G$ then $\langle \pi_1^{-1}, \pi_2^{-1}\rangle \in G$ and $\langle \pi_1 \circ \pi_1', \pi_2 \circ \pi_2'\rangle \in G$.

It is easy to check that given our definition of the action of a pair of permutations on higher types, a group of pairs of permutations of D_e and D_t is an equivalence generator (it is symmetric because it is closed under inverses, transitive because it is closed under composition, and reflexive because these conditions together imply that it includes the identity permutation).

Definition 34: A structure $M = \langle D, @, [\![\cdot]\!], v, G\rangle$ is a *permutation model* just in case $\langle D, @, [\![\cdot]\!], v\rangle$, is a standard applicative model, G is a group of pairs of permutations of D_e and D_t respectively, and M interprets identity by the equivalence generator $\{\sim_\pi\}_{\pi\in G}$.

Not every permutation model is a pure model or even a pristine model, because the equivalence generator may fail to be pure or pristine. Our task will be to construct examples which do satisfy these conditions.

C.2.1 Models with Opacity at Type *e*

In this section we will show how to give simple intensional models of Classical Purity, which exhibit opacity at type e. But we will also show that it is impossible to give intensional models of Classical Purity which exhibit opacity at type t: Classical Purity and $\approx$ Booleanism imply that there is no opacity at type t. This result will motivate slightly more complex models of Classical Purity that we will provide in the next subsection.

Definition 35: A structure $\langle E, W, [\![\cdot]\!], v, G\rangle$ is a *basic intensional permutation model* just in case $M = \langle E, W, [\![\cdot]\!], v\rangle$ is an intensional model, $G = G_e \times \{Id_{D_t}\}$ where G_e is a group of permutations of E, and $[\![=_\sigma]\!]^g = \lambda xz.\{w \in W : \pi_\sigma(x) = z \text{ for some } \pi \in G\}$.

Here the group of pairs of permutations is the cross-product of the group of permutations G_e of D_e with the singleton $\{Id_{D_t}\}$ which is a group of permutations of D_t. It's easy to check that this group of pairs of permutations is a pure equivalence generator; by definition, the model interprets identity by this equivalence generator. One can also readily check that Propositional Conjunction Congruence, Identity Congruence and Identifying Identities are valid on any basic intensional permutation model. Since $\approx$ Booleanism is valid on any intensional model, and in basic intensional permutation models $\approx^t$ coincides with $=_t$, these models also validate Booleanism. Finally, any full model in which G is nontrivial will exhibit opacity at type e; such models demonstrate that Classical Purity is consistent with opacity.

Interestingly, it turns out that in the simple intensional setting in which we have been working, it is not possible to use nontrivial permutations of D_t while still validating Propositional Conjunction Congruence:

Proposition 36: *$\approx$ Booleanism and Classical Purity imply that* $p =_t q \rightarrow (Fp \leftrightarrow Fq)$.

Proof: By $\approx$ Booleanism $p^t \vee \neg p^t \approx q^t \vee \neg q^t$, and also $q \vee \neg q \approx \top$ (where $\top$ abbreviates the pure term $(===) \vee \neg(===)$. Suppose for the remainder of the proof that $p = q$. Then $(p \vee \neg q = (p \vee \neg p)$ and therefore $(p \vee \neg q) = \top$ (by Propositional Conjunction Congruence and Purity). It follows that $(p \vee \neg q) \approx \top$ (by Lemma 1). By parallel reasoning, $q \vee \neg p \approx \top$.

By $\approx$ Booleanism, $q \approx (q \wedge \top)$ and hence by the claim just shown that $q \approx (q \wedge (p \vee \neg q))$. Moreover, by $\approx$ Booleanism, it follows from this that $q \approx ((q \wedge p) \vee (q \wedge \neg q))$ and thus $q \approx (q \wedge p)$. By parellel reasoning for p, $p \approx q \wedge p$. Hence $p \approx q$. Given Universal Instantiation, it follows that $Fp \leftrightarrow Fq$. □

C.2.2 Models with Opacity at Type t

To allow for opacity at type t consistently with Classical Purity, we must move to models which do not validate $\approx$ Booleanism. One can in fact give such models which validate Booleanism (which concerns identity rather than $\approx$). However, as noted in the main text, Booleanism is unattractive given Classical Purity. So we will here describe models in which propositions do not form a Boolean algebra but instead form an "agglomerative algebra" (Goodman (forthcoming)).

For simplicity we are going to work in a setting where the only base type is type t; this means we don't provide a domain D_e. While this simplifies the characterization of the models, nothing essential hangs on it; it is straightforward to generalize the construction.

Example 37: Given a complete atomic Boolean algebra $\mathcal{L} = \langle L, \wedge^*, \neg^*, \top \rangle$, we construct an applicative model $\langle D, @, [\![\cdot]\!], v \rangle$ as follows.

For the domains: we let $D_t = L \times N$, where $\mathcal{N} = \langle N, min(\cdot,\cdot), \omega \rangle$ and $N = \mathbb{Z} \cup \{\omega\}$. Note that $\mathcal{N}$ is a bounded meet-semilattice. We let $D_{\sigma\to\tau} = \{\langle f, n\rangle \in (D_\tau^{D_\sigma} \times N) : \text{proj}^2(f(x)) = min(n, \text{proj}^2(x)) \text{ for all } x \in D_\sigma\}$ (where $\text{proj}^i(\langle x_1, ..., x_n\rangle) = x_i$).

For application: we let $@ : \langle f, n\rangle \in D_{\sigma\to\tau}, x \in D_\sigma \mapsto f(x)$.

We use a group of automorphisms to give the interpretation of identity, as above. In particular, let G be the group of automorphisms of N that shift all elements of $\mathbb{Z}$ by the addition of a some element of $\mathbb{Z}$ and hold ω fixed. Given $h \in G$ we define an automorphism π_h on all types as follows: $\pi_h\langle p, n\rangle = \langle p, h(n)\rangle$ for all $\langle p, n\rangle \in D_t$ and for $\langle f, n\rangle \in D_{\sigma\to\tau}$, $\pi_h(\langle f, n\rangle) = \langle \pi_h \circ f \circ \pi_h^{-1}, h(n)\rangle$.

$[\![\cdot]\!]$ is such that:

- $[\![\lambda x.F]\!]^g = \langle \lambda z.[\![F]\!]^{g(z/x)}, n\rangle$, where n is the greatest lower bound of the second coordinates of the constants and free variables in $\lambda x.F$, relative to g.
- $[\![\neg]\!]^g = \langle f, \omega\rangle$, where $f : \langle l, n\rangle \in D_t \mapsto \langle \neg^* l, n\rangle$.
- $[\![\wedge]\!]^g = \langle f, \omega\rangle$, where $f : \langle l, n\rangle, \langle l', m\rangle \in D_t \mapsto \langle l \wedge^* l, min(n, m)\rangle$.
- $[\![\forall_\sigma]\!]^g = \langle f, \omega\rangle$, where $f : \langle l, n\rangle \in D_{\sigma\to t} \mapsto \langle \bigwedge^* \{\text{proj}^1(l(x)) : x \in D_\sigma\}, n\rangle$.
- $[\![=_\sigma]\!]^g = \langle f, \omega\rangle$, where $f : \langle l, n\rangle \in D_\sigma \mapsto \langle f_l, n\rangle$, and $f_l : \langle k, m\rangle \in D_\sigma \mapsto \langle \top, min(n, m)\rangle$, if $\langle l, n\rangle \sim_\sigma \langle k, m\rangle$, and $f_l : \langle k, m\rangle \in D_\sigma \mapsto \langle \bot, min(n, m)\rangle$, if $\langle l, n\rangle \nsim_\sigma \langle k, m\rangle$.

$v(\cdot)$ is such that, for some atom a of $\mathcal{L}$, $v(\phi) = 1$ just in case $a \wedge^* \text{proj}^1(\phi) = a$.

It is easy to check that every such model is standard and that $\{\sim_{\pi_h}\}_{h\in G}$ is a pure equivalence generator. Moreover, Identity Congruence and Propositional Conjunction Congruence are valid on every such model. The models, moreover, clearly exhibit opacity at type t. Thus Classical Purity is consistent with opacity at type t as well as at type e.

Appendix D: Models of Free Opacity

To give models of Free Applicativism, we first define the notion of super-transparency.

Definition 38: Fix an applicative model $\langle D, @, [\![\cdot]\!], v\rangle$. For any type $\sigma \rightarrow \tau$, $f \in D_{\sigma\rightarrow\tau}$ is *transparent* if and only if for every $y, z \in D_\sigma$ if $v(([\![=_\sigma]\!]@y)@z) = 1$ then $v(([\![=_\tau]\!]@(f@y))@(f@z)) = 1$.

For any $\rho \in \mathcal{T}$, $x \in D_\rho$ is *super-transparent* if and only if $\rho \in \{e, t\}$, or $\rho = \sigma \rightarrow \tau$, f is transparent, and for any super-transparent $z \in D_\sigma$, $f@z$ is super-transparent.

Given an applicative model M, we write ${}^{\mathbf{M}}\mathbf{S}_\sigma$ for the set of super-transparent elements of D_σ.

Definition 39: A structure $\mathcal{M} = \langle E, W, [\![\cdot]\!], v, \sim\rangle$ is a *free congruential model* just in case $[\![\forall_\sigma]\!]^g = \lambda f : f \in D_{\sigma\rightarrow t}.\bigcap_{x\in{}^{\mathbf{M}}\mathbf{S}_\sigma} fx$, and otherwise everything is exactly as in the definition of a congruential intensional model.

Free congruential models fail to be intensional models because the clause for the quantifiers does not range over the whole of D_σ, but instead ranges over a subset of D_σ, ${}^{\mathbf{M}}\mathbf{S}_\sigma$. It is readily verified that Free Applicativism is valid on these models. In particular, the validity of Free Substitution and Existence Preservation follows straightforwardly from the restriction of the quantifiers to the super-transparent entities. One can also easily check that Constant Existence and Continued Existence are valid on them.

We can prove the consistency of Purity with Free Instantiation using permutation models, but it turns out that in these models we cannot prove the consistency of Purity, Free Instantiation, and Existence Preservation.

Definition 40: Fix a permutation model $M = \langle D, @, [\![\cdot]\!], v, G\rangle$. An element $x \in D_\sigma$ is *fixed* just in case $\sigma \in \{e, t\}$ or for all $\pi \in G$, $\pi_\sigma(x) = x$.

Given a model M, we write the set of fixed elements of D_σ as ${}^{M}\mathbf{F}_\sigma$.

Definition 41: A structure $M = \langle E, W, [\![\cdot]\!], v, G\rangle$ is a *basic pure free model* just in case $[\![\forall_\sigma]\!]^g = \lambda f : f \in D_{\sigma\rightarrow t}.\bigcap_{x\in{}^{M}\mathbf{F}_\sigma} fx$, and everything else is as in a basic pure intensional model.

These models validate Purity and Free Instantiation, together with e/t-Existence and Pure Existence. But the models cannot be extended to validate Existence Preservation. Permutation models exhibit opacity at any types only if the permutations at e or t are nontrivial. But if these permutations are nontrivial, then Application Equivalence must fail for expressions of type $e \rightarrow t$ or $t \rightarrow t$. However we can show that Free Purity implies Application Equivalence for these types:

Proposition 42: *Free Purity implies e/t-Application Equivalence, i.e. $F = G \rightarrow (Fa^\sigma \leftrightarrow Ga^\sigma)$ for $\sigma \in \{e, t\}$.*

Proof: We give the proof for the case of e; the proof for t is exactly parallel. (i) $\exists X(X = (\lambda y\lambda X.Xy))$ (Pure Existence); (ii) $\exists x(x =_e a)$ (e/t-Existence); (iii) $\exists X(X = (\lambda X.Xa))$ (Existence Preservation); (iv) $F = G \rightarrow \forall X(XF \leftrightarrow XG)$

(Leibniz's Law); (v) $F = G \rightarrow ((\lambda X.Xa)F \leftrightarrow (\lambda X.Xa)G)$ (Free Instantiation, (iii)); (vi) $F = G \rightarrow (Fa \leftrightarrow Ga)$ (Beta-Eta Equivalence). □

This Proposition shows that permutation models cannot be straightforwardly altered to validate Existence Preservation consistently with our other assumptions. But Kripke intensional structures can. This may seem a little surprising. We showed earlier (Example 31) that $\forall_{e\rightarrow t}$ fails to be transparent in simple models of this kind. But first, while the quantifiers of this and higher types may fail to be transparent, $\forall_e$ and $\forall_t$ are in fact transparent. Second, if we restrict the domains of higher quantifiers appropriately, even those higher quantifiers will be transparent, as we will now show.

Definition 43: Fix a Kripke intensional structure $A = \langle E, W, V, \sim \rangle$, with domains D defined as usual. $f \in D_\sigma$ is *nice* just in case $f \sim f$.

Given such a structure A, we write the set of nice entities of type σ as ${}^A\mathbf{N}_\sigma$.

Definition 44: A structure $\langle E, W, V, \sim, [\![\cdot]\!], v\rangle$ is a *free pure Kripke model* just in case $A = \langle E, W, V, \sim \rangle$ is a Kripke intensional structure, $[\![\forall_\sigma]\!]^g = \lambda z : f \in D_{\sigma\rightarrow t} . \bigcap_{x\in {}^A\mathbf{N}_\sigma} fx$, and everything else is as in a Kripke pristine intensional model.

Free pure Kripke models clearly exist. They validate e/t-Existence, Free Instantiation, Booleanism and $\approx$-Booleanism. The way that logical relations are defined guarantees that the nice entities are closed under application: if $f \sim f$, then for any x such that $x \sim x$, $fx \sim fx$. This fact, together with the semantics for the quantifiers, guarantees that such models validate Existence Preservation. We showed earlier that if something is nice, it is transparent. Given this fact, it is easy to see that free pure Kripke models validate Leibniz's Law and Free Substitution. Given the altered semantics for the quantifier, free pure Kripke models (unlike Kripke pristine intensional models) do not validate Functionality, though they do validate $\forall X \forall Y(\forall x(Xx = Yx) \rightarrow X = Y)$.[56] But the altered semantics for the quantifiers means the models now validate Quantified Application Congruence (in spite of still failing to validate Application Congruence, which they must since they validate Pristineness).

We observed earlier that the denotations of combinators, identity, conjunction, and negation are all nice in Kripke pristine intensional models. Free pure Kripke models are the same as pristine intensional models with respect to the definition of $\sim$ and the interpretation of these operations. So free pure Kripke models also validate Pristineness. But in free pure Kripke models we can also show that for all σ, $[\![\forall_\sigma]\!]^g \sim [\![\forall_\sigma]\!]^g$. To see this suppose $f \sim g$, i.e. for all x, y so that $x \sim y$, $fx \sim gy$. In particular, for all x such that $x \sim x$, i.e. $x \in {}^A\mathbf{N}_\sigma$, $fx \sim gx$. Now, the definition of $\sim$ on type t guarantees that $\sim$ is a congruence with respect to arbitrary intersection, and this means that if for all $x \in {}^A\mathbf{N}_\sigma$, $fx \sim gx$ then $\bigcap_{x\in {}^A\mathbf{N}_\sigma} fx \sim \bigcap_{x\in x\in {}^A\mathbf{N}_\sigma} gx$. But then it follows that the denotations of the

[56] Suppose F and G are not transparent. They may have the same functional behavior on all nice entities (and so satisfy the antecedent of Functionality), but fail to be identical (because they produce different values when they have opaque entities as arguments), so $\forall x(Fx = Gx) \rightarrow F = G$ may fail. This kind of counterexample can only arise for F and G which are not nice, so the quantified principle in the main text holds.

quantifiers are also nice, as required. This guarantees that the new models validate not just Pristineness but also Purity.

Finally, we show that the model also validates Orthodoxy. First a definition, and a preliminary result:

Definition 45: Fix an applicative model $M = \langle D, @, [\![\cdot]\!], v\rangle$. $M' = \langle D', @', [\![\cdot]\!]', v'\rangle$ is a *submodel of M* just in case it is an applicative model and

- for all σ, $D'_\sigma \subseteq D_\sigma$;
- $@' = @|_{D'_\sigma}$;
- $[\![\cdot]\!]^g = [\![\cdot]\!]'^g$ for all g on which the latter is defined;
- $v' = v|_{D'_t}$.

The following result is immediate given the definition of a submodel:

Theorem 46: *Fix an applicative model M and a submodel of M, M′. If φ is a closed formula which is valid on M′, then φ is valid on M.*

We now characterize a class of structures which we will show to be submodels of free pure Kripke models:

Definition 47: Fix a free pure Kripke model $M = \langle E, W, V, \sim, [\![\cdot]\!], v\rangle$, with domains D defined as usual. For $A = \langle E, W, V, \sim\rangle$, the *pared model* of M is the structure $\langle {}^A\mathbf{N}, Id_D, [\![\cdot]\!]', v'\rangle$, where $[\![\cdot]\!]^g = [\![\cdot]\!]'^g$ for all g with range in ${}^A\mathbf{N}$, and $v' = v|_{D'_t}$.

Proposition 48: *Fix a free pure Kripke model M and a logical signature* Σ *containing only the constants* $\wedge, \neg, \forall_\sigma$ *and* $=_\sigma$. *The pared model of M is a submodel of M which validates Orthodoxy.*

To show that the pared model is a submodel, the only non-obvious claim is that the pared model is in fact an applicative model. But this claim follows straightforwardly from the fact that the denotations of combinators and the constants of the language are all nice. The pared model is also a standard model, so it validates Orthodoxy. (In fact it also validates Functionality.) It follows from the Proposition and Theorem 46 that free pure Kripke models also validate Orthodoxy.

Appendix E. Glossary of Principles

Principles are listed in alphabetic order, followed by the section in which they first appear, in parentheses.

Application Congruence $F = G \rightarrow Fa = Ga$ (Section 2)

Application Equivalence $F = G \rightarrow (Fa \leftrightarrow Ga)$ (Section 3.1)

Applicative Individuation $\forall x(Fx = Gx) \leftrightarrow F = G$ (Section 3.1)

Beta-Eta Equivalence $\varphi \leftrightarrow \psi$, provided φ and ψ are $\beta\eta$-equivalent (Section 2)

Booleanism $a = b$, whenever $a \leftrightarrow b$ is a theorem of propositional logic (Section 3.1)

Conjunction Congruence $(a = b \wedge c = d) \rightarrow (a \wedge c) = (b \wedge d)$ (Section 3.1)

Constant Existence $\exists X(X = O)$ if O is an extended logical constant (n. 45)

Constant Transparency $a = b \rightarrow Oa = Ob$, if O is an extended logical constant (Section 3.1)

Continued Existence $\exists X(X = O)a$ if O is an extended logical constant (n. 45)

e-**Atomic Substitution** $a^e = b^e \rightarrow (Fa \leftrightarrow Fb)$ (Section 2)

e/*t*-**Existence** $\exists x(x = \alpha)$ if α is of type e or t (Section 4)

Equivalence $a = a \wedge (a = b \wedge a = c \rightarrow b = c)$ (Section 2)

Existence Preservation $\forall X \forall x \exists Z(Z = Xx)$ (Section 4)

Free Instantiation $\exists x(x = a) \rightarrow \forall x\varphi \rightarrow \varphi[a//x]$ (Section 4)

Free Substitution $\exists X(X = F) \rightarrow a = b \rightarrow Fa = Fb$ (Section 4)

Functionality $\forall x(Fx = Gx) \rightarrow F = G$ (Section 3.1)

Generalization If $\varphi \rightarrow \psi \in T$, then $\varphi \rightarrow \forall x\psi \in T$, where x does not occur free in φ (Section 2)

Generalized Leibniz's Law $a = b \rightarrow \forall X \forall x((Xx)a \leftrightarrow (Xx)b)$ (Section 4)

Identifying Identities $(a = a) = (b = b)$ (Section 3.2)

Identity Congruence $(a = b \wedge c = d) \rightarrow (a = c) = (b = d)$ (Section 3.1)

Instantiation Equivalence $(\lambda X.Xa)F \leftrightarrow Fa$ (Section 2)

LE Substitution $a \approx b \rightarrow (\varphi \leftrightarrow \varphi[b/a])$ (Section 4)

Leibniz's Law $a = b \rightarrow \forall X(Xa \leftrightarrow Xb)$ (Section 1)

Lift Congruence $a = b \rightarrow (\lambda X.Xa) = (\lambda X.Xb)$ (Section 2)

Material Equivalence $p = q \rightarrow (p \leftrightarrow q)$ (Section 2)

Pristineness $a = b \rightarrow Fa = Fb$, provided F is a pristine term (Section 3.2)

Propositional Conjunction Congruence $(p =_t p' \wedge q =_t q') \rightarrow (p \wedge q =_t p' \wedge q')$ (Section 3.2)

Pure Existence $\exists x(x = a)$, provided a is pure (n. 48)

Pure Instantiation $\forall x\varphi \rightarrow \varphi[a//x]$, provided a is a pure term (n. 54)

Pure Universal Distinctness $(\lambda xyZ.(x = y) \rightarrow (Zx \leftrightarrow Zy)) \neq \alpha$ for every pure α such that $\alpha xyZ \in T$ for x,y,Z of appropriate types (n. 54)

Purity $a = b \rightarrow Fa = Fb$, provided F is a pure term (Section 3.2)

Quantified Application Congruence $F = G \rightarrow \forall x(Fx = Gx)$ (Section 3.1)

Quantified Substitution $\forall x\forall y(x = y \rightarrow (\varphi \leftrightarrow \varphi[y/x]))$ (Section 1)

Reduction Congruence $(\lambda x.\varphi)a = (\lambda x.\psi)b \rightarrow \varphi[a/x] = \psi[b/x]$ (Section 2)

Strong Transparency Preservation $a = b \rightarrow (((Oc_1)\ldots c_n)a = ((Oc_1)\ldots c_n)b)$, for O an extended logical constant of the enriched language (n. 28)

Substitution $a = b \rightarrow (\varphi \leftrightarrow \varphi[b/a])$ (Section 1)

t-Atomic Substitution $a^t = b^t \rightarrow (Fa \leftrightarrow Fb)$ (Section 2)

Transparency Preservation $a = b \rightarrow (Oc)a = (Oc)b$, if O is an extended logical constant (Section 3.1)

Universal Instantiation $\forall x\varphi \rightarrow \varphi[a//x]$, where a is free for x in φ (Section 3)