Logic and **L**ogical **P**hilosophy
Volume 28 (2019), 661–681
DOI: 10.12775/LLP.2019.017

Stephen Pollard

LOGIC IN THE LAND OF MAKE-BELIEVE

Abstract. Philosophers call it "contagion" when pretense influences belief, behavior, perception, or emotion. This pejorative terminology is justified in some cases: fantasy and imagination *can* exercise a pathological influence. This essay, however, reviews some logical techniques that allow pretense to govern belief in a rational and beneficial way. Philosophers might want similar techniques in their tool-kits when they explore interactions between belief and pretense.

Keywords: abstraction; interpretation; contagion; pretense

1. Introduction

Belief can influence pretense and pretense can influence belief. The latter relationship (the modification of belief by imagination) has acquired the label "contagion" in the philosophical literature — suggesting that some philosophers view it with suspicion. This negative attitude makes some sense given the multitude of bad ways for our fantasies to rule our beliefs. Here, however, we will focus on techniques that give pretense a useful role in the rational fixation of belief. We will review some logical tricks and devices that deserve a place in the philosopher's tool kit because they allow for self-conscious pretenses that are not just harmless but fruitful.

2. The same list

Suppose you take your pen and your piece of paper and make a list. I take my pen and my piece of paper and make a list. We compare the results and are surprised to find that our lists are the same.

Your list = My list.

Received May 7, 2018. *Revised* October 28, 2018. *Published online* June 11, 2019

What makes us think they are the same? Simple: they list the same things. Lists that list the same things are the same.[1] Yet they are *not* the same: they are not really *identical*. If you burn your list, it does not follow that you have burned my list. If your list is in cursive, it does not follow that my list is in cursive. This "same list" business is a pretense that can miscarry in two ways: it can unravel internally and it can saddle us with falsehoods when we drop the pretense. If our pretense is to remain coherent and if the results we reach inside our pretense are to stand up when we drop the pretense, we have to monitor the propositions we affirm and the inferences we make on the basis of our pretended identities. In particular, we have to avoid contexts that let us distinguish between lists that have the same entries.[2]

Let $\approx$ be the relation "having the same entries": if X and Y are lists, then $X \approx Y$ if and only if X and Y list the same things. Let us say that a formula of the form $\alpha = \beta$ is an EQUATION while $\alpha \approx \beta$ is the corresponding EQUIVALENCE. Within our pretense we replace sincere equivalences with pretended equations. (We say that X and Y are the same list when we are really only justified in saying that they have the same entries.) When we drop our pretense, we reverse the process: replacing equations with equivalences. Within our pretense, we might assert various instances of Leibniz's Law — that is, propositions of the form

$$\forall\bar{\alpha}\forall X, Y((\phi(X) \wedge X = Y) \rightarrow \phi(Y)) \qquad (LL=)$$

where 'X' and 'Y' are list-variables (variables ranging over lists) and $\forall\bar{\alpha}$ is a string of universal quantifiers of appropriate type binding any further

[1] At least, this is the rule for lists that involve no ranking or ordering and in which multiple listings of the same entry count the same as a single listing. From here on, we will only consider lists that obey this rule.

[2] There is a substantial literature on this style of logical make-believe. Here is just a taste. Moritz Pasch treats our pretense as a form of IMPLICIT DEFINITION. See, for example, [12]; translated into English as [13]. For a general discussion of Pasch's views on mathematical pretenses, see [17]. Hermann Weyl and Paul Lorenzen describe our pretense as an application of DEFINITION BY ABSTRACTION. See [23, pp. 8–13]; [10, pp. 105–111]; [9]; and [15]. The idea is that equivalence relations can better impersonate identity when we ABSTRACT FROM (that is, when we systematically avoid) contexts that distinguish between equivalent objects. What we call a *pretense*, Lorenzen calls a *façon de parler*. Cf. Quine's remark that "the metaphorical use of the identity sign for what is really not identity" is "a manner of speaking" [19, p. 118]. For an especially clear discussion of this manner of speaking see [3, pp. 159–161]. For some criticisms of Lorenzen, see [21, pp. 26–28]. For contributions by Ignacio Angelelli and many others, see the references in [1]. For some insight into the historical background, see [11].

variables free in the formula ϕ. When we drop the pretense, equations become equivalences and, so, instead of ($LL =$) we have

$$\forall\bar{\alpha}\forall X, Y((\phi^*(X) \wedge X \approx Y) \rightarrow \phi^*(Y)) \qquad (LL \approx)$$

where ϕ^* is the result of replacing each equation in ϕ with the corresponding equivalence. We might insist that this is what we *really* mean when we assert instances of ($LL =$) within our pretense. At the very least, we want to be justified in asserting an instance of ($LL \approx$) whenever our pretense leads us to affirm the corresponding instance of ($LL =$). So we need to be careful to assert only those instances that translate into truths when we drop our pretense.

Let us consider an example. Would it be safe, inside our pretense, to affirm the following?

$$\forall X, Y, Z((X = Z \wedge X = Y) \rightarrow Y = Z).$$

Would this preserve the coherence of our pretense and would it be consistent with the project of using make-believe as a guide to truth? According to our standard, the answer is "yes." To see why, first note that if $\phi(X)$ is '$X = Z$' and $\phi(Y)$ is '$Y = Z$', then $\phi^*(X)$ is '$X \approx Z$' and $\phi^*(Y)$ is '$Y \approx Z$'. So our translation is:

$$\forall X, Y, Z((X \approx Z \wedge X \approx Y) \rightarrow Y \approx Z).$$

This is a true statement about lists. If X has the same entries as both Z and Y, then Y has the same entries as Z (since "sameness of entries" is symmetric and transitive). So we can safely affirm

$$\forall X, Y, Z((X = Z \wedge X = Y) \rightarrow Y = Z)$$

within our pretense: this proposition is a reliable guide to truth outside our pretense. If premises of the form $\alpha = \gamma$ and $\alpha = \beta$ turn out true when we drop our pretense (that is, if $\alpha \approx \gamma$ and $\alpha \approx \beta$ are true), then $\beta = \gamma$ will also turn out true when we drop our pretense (that is, $\beta \approx \gamma$ will be true).

For another case where things work out well, we turn to the "listing" relation: the relation between a list and its entries. X LISTS z if and only if z is an entry of X. Consider the proposition

$$\forall z\forall X, Y((X \text{ lists } z \wedge X = Y) \rightarrow Y \text{ lists } z)$$

where 'z' is an individual-variable (a variable ranging over non-lists). To see how to apply our interpretation technique here, first note that if $\phi(X)$

is 'X lists z', then $\phi^*(X)$ is still 'X lists z'. When a formula harbors no equations, either implicitly or explicitly, the operation of replacing equations with equivalences does not actually change anything. We end up with the same formula we had at the start. So our translation is

$$\forall z \forall X, Y((X \text{ lists } z \wedge X \approx Y) \rightarrow Y \text{ lists } z).$$

This is another true statement about lists. If X and Y have the same entries, then Y will list everything X does. So, inside our pretense, we can confidently affirm that

$$\forall z \forall X, Y((X \text{ lists } z \wedge X = Y) \rightarrow Y \text{ lists } z).$$

This affirmation will not render our pretense incoherent and it will be a reliable guide to truth outside our pretense.

Things do not always work out so well. For example, we must not affirm this instance of Leibniz's Law:

$$\forall X, Y((X \text{ burned up} \wedge X = Y) \rightarrow Y \text{ burned up}).$$

Here our translation is the following falsehood:

$$\forall X, Y((X \text{ burned up} \wedge X \approx Y) \rightarrow Y \text{ burned up}).$$

If, for example, your list burned up and your list has the same entries as mine, it does not follow that my list burned up. You can incinerate a list without incinerating all equivalent lists. So inferences of the form

$$(\alpha \text{ burned up} \wedge \alpha = \beta) \implies \beta \text{ burned up}$$

can make our pretense incoherent and can saddle us with absurdities when we drop the pretense. Suppose we acknowledge that your list burned up, but mine did not. Suppose, calling on our feigned belief in the identity of our lists, we infer that my list burned up. Our position will then be that my list both did and did not burn up. Somehow or other, we must keep clear of this trap.

One solution would be to avoid premises (such as, "Your list burned up") that get us into trouble. For starters, we could make our current topic, incineration, entirely taboo. That, however, would be going too far. It is the incineration of *lists* that gets us in trouble. We might still want to make inferences about the incineration of list-entries that are not themselves lists. Such inferences can be entirely innocent. For example, if you made a list of manuscripts, one of which burned up,

and my list has the same entries as your list, then it really does follow that an entry of my list burned up. We will not necessarily want to ban that inference. Furthermore, a general taboo on talk about incineration (and, to mention just a few other topics, spatial location, surface area, reflectivity, odor, ink color, aesthetic value) would not really solve our problem. There is a topic central to discourse about lists that will still get us into trouble.

That topic is our old friend the "listing" relation. We already showed that the proposition

$$\forall z \forall X, Y((X \text{ lists } z \wedge X = Y) \rightarrow Y \text{ lists } z)$$

is safe: we are able to interpret it as a truism about lists. Now, however, consider this:

$$\forall X, Y, Z((Z \text{ lists } X \wedge X = Y) \rightarrow Z \text{ lists } Y).$$

Our translation is the following falsehood:

$$\forall X, Y, Z((Z \text{ lists } X \wedge X \approx Y) \rightarrow Z \text{ lists } Y).$$

This is false because a list can list a list without listing every list that happens to have the same entries. A list of your lists would, in fact, list *your* lists. My list may have the same entries as one of yours, but it is still *my* list and, so, does not belong on a list of your lists. At least, this is how matters stand when we have dropped our pretense and are carefully distinguishing between '$=$' and '$\approx$'. When faced with sober senses, pretense-driven inferences of the form

$$(\gamma \text{ lists } \alpha \wedge \alpha = \beta) \implies \gamma \text{ lists } \beta$$

turn out to be unreliable. They can lead us to make false claims, not about side issues like mass or color, but about the central concern in discourse about lists: the listing relation itself [14].

3. Type-1 lists

Well what are we to do? We cannot ban talk about what-lists-what because that would leave us with almost nothing of real interest to say about lists. It is fortunate that we have less extreme options. We can

pinpoint the source of our trouble: lists that appear as entries of lists. If list X is an entry of list Z and X has the same entries as Y it does not follow that Z lists Y. $\approx$'s impersonation of identity can break down here. We could avoid this problem by refusing to talk about lists that have lists as entries. Say that a list is TYPE-1 if none of its entries are lists. Suppose we agree to avoid any reference to lists that are not type-1. In this setting, when we affirm that

$$\forall X, Y, Z((Z \text{ lists } X \wedge X = Y) \rightarrow Z \text{ lists } Y)$$

it is with the understanding that the lists X, Y, Z will be type-1. We make this understanding explicit in a new interpretation

$$\forall X, Y, Z \in T_1((Z \text{ lists } X \wedge X \approx Y) \rightarrow Z \text{ lists } Y)$$

where '$\forall X, Y, Z \in T_1$' is a universal quantifier to be read as "for all type-1 lists X, Y, Z." (You might think of T_1 as the class of all type-1 lists.) If X is a list and Z is a list that lists no lists, then Z will not list X and, hence, a conditional whose antecedent asserts that Z does list X will be vacuously true. That is, we can now approve of the proposition

$$\forall X, Y, Z((Z \text{ lists } X \wedge X = Y) \rightarrow Z \text{ lists } Y)$$

because we have figured out how to interpret it as a true (albeit rather uninteresting) statement about lists.

We need to give a more general account of our translation scheme. Our interpretation of a formula ϕ is now ϕ_1^* where the latter is the result of restricting any list-quantifiers in ϕ^* to T_1: for example, translating '$\forall Z$' as '$\forall Z \in T_1$' and '$\exists Z$' as '$\exists Z \in T_1$'. We now interpret instances of Leibniz's Law as

$$\forall \bar{\alpha} \forall \bar{\beta} \in T_1 \forall X, Y \in T_1((\phi_1^*(X) \wedge X \approx Y) \rightarrow \phi_1^*(Y))$$

where $\forall \bar{\alpha}$ is a string of univeral quantifiers binding any individual-variables that would otherwise be free and $\forall \bar{\beta} \in T_1$ is a string of universal quantifiers doing the same for list-variables. Consider this instance of Leibniz's Law:

$$\forall X, Y((\exists Z(Z \text{ lists } X) \wedge X = Y) \rightarrow \exists Z(Z \text{ lists } Y)).$$

That is, if X is an entry of some list and X is the same as Y, then Y is an entry of some list. This looks like the sort of thing that could get

us in trouble. After all, an unlisted list can have the same entries as a listed one. However, our new interpretation is:

$$\forall X, Y \in T_1((\exists Z \in T_1(Z \text{ lists } X) \wedge X \approx Y) \rightarrow \exists Z \in T_1(Z \text{ lists } Y)).$$

If X is a list, then there cannot be a list Z that lists no lists and yet lists X. So our translation is vacuously true. This shows that the translated proposition is safe: we are able to interpret it as a true (albeit uninteresting) claim about lists.

4. Hereditarily invariant lists

That solves our problem with the listing relation, but the price is stiff: a substantial limitation on the power and interest of our talk about lists. There is an alternative. Instead of avoiding all lists that list lists, we could avoid the ones that get us in trouble: lists that list lists without listing all equivalent lists. That is, if we are to discuss a list Z, we will want it to pass the following test of INVARIANCE:

$$\forall X, Y((Z \text{ lists } X \wedge X \approx Y) \rightarrow Z \text{ lists } Y).$$

If list X is an entry of invariant list Z, then Z will list every list equivalent to X. If, within our pretense, we discuss only invariant lists, then the problematic claim

$$\forall X, Y, Z((Z \text{ lists } X \wedge X = Y) \rightarrow Z \text{ lists } Y)$$

will translate as

$$\forall X, Y, Z \in I((Z \text{ lists } X \wedge X \approx Y) \rightarrow Z \text{ lists } Y)$$

where '$\forall X, Y, Z \in I$' is a universal quantifier to be read as "for all invariant lists X, Y, Z." That is, we will interpret the problematic claim as a true statement about invariant lists. An *invariant* list that lists X will, by the definition of invariance, list every list equivalent to X. Our other problematic claim

$$\forall X, Y((\exists Z(Z \text{ lists } X) \wedge X = Y) \rightarrow \exists Z(Z \text{ lists } Y))$$

will translate as the true statement

$$\forall X, Y \in I((\exists Z \in I(Z \text{ lists } X) \wedge X \approx Y) \rightarrow \exists Z \in I(Z \text{ lists } Y)).$$

If X is an entry of an invariant list, then every list equivalent to X will be an entry of an invariant list. We seem to have hit upon a successful strat-

egy: interpret each list-quantifier as a quantifier restricted to I just as, above, we interpreted each list-quantifier as a quantifier restricted to T_1.

Unfortunately, we soon run into a new problem. At the core of our pretense is the principle that lists with the same entries are the same:

$$\forall X, Y((\forall z(X \text{ lists } z \leftrightarrow Y \text{ lists } z) \wedge \forall Z(X \text{ lists } Z \leftrightarrow Y \text{ lists } Z)) \rightarrow X = Y).$$

When we restrict each list-quantifier to I, we obtain:

$$\forall X, Y \in I((\forall z(X \text{ lists } z \leftrightarrow Y \text{ lists } z) \wedge \forall Z \in I(X \text{ lists } Z \leftrightarrow Y \text{ lists } Z)) \rightarrow X \approx Y).$$

This says that if X and Y list the same individuals and the same *invariant* lists, then X and Y have the same entries. But that is not right. X and Y could agree on individuals and invariant lists, but disagree on non-invariant lists. Although our formula stipulates that X and Y are invariant, it does not require that their entries be invariant. So X could have a non-invariant entry that Y lacks.

We might decide, then, to focus on invariant lists whose entries are all either individuals or invariant lists. Say that such lists are INVARIANT$^+$. Our interpretation of our core principle will now be: if invariant$^+$ lists X and Y list the same individuals and the same invariant$^+$ lists, then X and Y have the same entries. Sadly, this gets us nowhere. Since X and Y are invariant$^+$, any lists that appear among their entries will be invariant; but there is no guarantee that those entries will be invariant$^+$. So X and Y could agree on individuals and invariant$^+$ lists while disagreeing on lists that are merely invariant. It would be just as futile to focus on invariant^{++} lists: that is, invariant$^+$ lists whose entries are all either individuals or invariant$^+$ lists. The problem is that invariant lists can have entries that are not invariant; invariant$^+$ lists can have entries that are not invariant$^+$; invariant^{++} lists can have entries that are not invariant^{++}; and so on. We need some sort of super-invariance property that is passed from lists to entries. Super-invariant lists really would have the same entries when they agree on individuals and super-invariant lists: since each list appearing as an entry would be super-invariant, agreement on super-invariant lists would mean agreement on lists in general.

Now for some good news: the property we need is ready and waiting. It is known as "hereditary invariance" [20, p. 118]. A list is HEREDITARILY INVARIANT if and only if it is invariant, its list-entries (its entries that

are lists) are invariant, the list-entries of its list-entries are invariant, the list-entries of the list-entries of its list-entries are invariant, and so on. If the "and so on" makes you uneasy, rest assured that we could use a trick (Frege's definition of the ancestral) to eliminate it.[3] We shall take it for granted that hereditary invariance has been properly defined. The core principle of our pretense will now have the following interpretation

$$\forall X, Y \in H((\forall z(X \text{ lists } z \leftrightarrow Y \text{ lists } z) \wedge \forall Z \in H(X \text{ lists } Z \leftrightarrow Y \text{ lists } Z)) \rightarrow X \approx Y)$$

where '$\forall X, Y \in H$' is a universal quantifier to be read as "for all hereditarily invariant lists X, Y." Suppose X and Y are hereditarily invariant lists that list the same individuals and the same hereditarily invariant lists. Any list that appears as an entry of X or Y will be hereditarily invariant. (Entries of hereditarily invariant lists are all either individuals or hereditarily invariant lists.) So, in fact, X and Y will list the same individuals and the same lists. That is, X and Y will have the same entries. The core principle of our pretense comes out true under our new interpretation.

5. How to do it

We can now see how $\approx$ might carry off a successful impersonation of identity. Here success mean, first, that our pretense is internally consistent and, second, that we have a systematic way to interpret make-believe results as straightforward truths about lists. We start by addressing a problem we left unresolved. Recall that the predicate 'burned up' fails an invariance test.[4] That is, the following proposition is false:

$$\forall X, Y((X \text{ burned up} \wedge X \approx Y) \rightarrow Y \text{ burned up}).$$

[3] For one example of how to do so, see [18, p. 61]

[4] A *list* Z is invariant when

$$\forall X, Y((Z \text{ lists } X \wedge X \approx Y) \rightarrow Z \text{ lists } Y).$$

A *predicate* ϕ is invariant when

$$\forall X, Y((\phi(X) \wedge X \approx Y) \rightarrow \phi(Y)).$$

When a predicate expresses a property, we say that the property is invariant if and only if the predicate is [23, p. 9]. The property of being listed by Z is invariant if and only if Z is invariant.

A direct solution is to enforce a type restriction that prevents list-terms (such as list-variables) from occupying the subject position in the matrix

_____ burned up.

This will leave us unable to affirm that a list burned up. It will also leave us unable to *deny* that a list burned up, but that is as it should be. The predicate 'did not burn up' also fails the invariance test since the following proposition is false:

$$\forall X, Y((X \text{ did not burn up} \wedge X \approx Y) \rightarrow Y \text{ did not burn up}).$$

Even if your list, with the same entries as mine, avoided incineration, it does not follow that my list was as lucky. Once we impose our type-discipline, a list will just not be the sort of thing that either burns or fails to burn, much as the equator is not the sort of thing that is either heavy or light.[5] We will treat the formula

$$\forall X(X \text{ burned up} \vee X \text{ did not burn up})$$

as ungrammatical. We can still talk about the incineration of *individuals* and we can still accept

$$\forall x(x \text{ burned up} \vee x \text{ did not burn up})$$

as a logical truth. We deal similarly with other non-invariant predicates (such as 'is beautiful' or 'is in cursive') with the one crucial exception of the listing relation itself. It would be crippling to ban talk about what lists list. Instead, we can allow such talk, but limit it to a special class of lists. Limiting our list-talk to hereditarily invariant lists would have the advantage of letting lists be entries of lists. However, if this strategy proves unworkable in some context, we can just confine ourselves to type-1 lists (which, by the way, are all hereditarily invariant).

Let us consider an example. Suppose we are discussing some universe U of lists and are pretending that lists with the same entries are identical. We ask: *is there a list that lists exactly one list*? If we are safeguarding

[5] Imaginary scenarios are, notoriously, INCOMPLETE. If there is an imaginary spill in a pretend tea party, "there may be no fact of the matter (in the pretense) just how much tea spilled" [5, p. 25]. In our pretense about lists, there are no "facts of the matter" to support claims of incineration or non-incineration. It is not just that we have failed to provide for such facts: we have kept our pretense coherent by actively excluding them.

our pretense by recognizing only type-1 lists, then the answer is clear: no list lists exactly one list because no list lists any list at all. Suppose our outlook is broader: we are willing to consider any hereditarily invariant lists in U. What is our answer then? We first expose the logical structure of our question. To say that some list lists exactly one list is to say:

$$\exists X, Y \forall Z (X \text{ lists } Z \leftrightarrow Z = Y).$$

When we drop our pretense, we interpret this as

$$\exists X, Y \in H \forall Z \in H (X \text{ lists } Z \leftrightarrow Z \approx Y)$$

where H is the class of hereditarily invariant lists in U. It is easy to imagine a situation in which this is true. Suppose U features a type-1 list Y and a list X whose entries are exactly the lists in U equivalent to Y. Then both X and Y will be hereditarily invariant and, furthermore, hereditarily invariant lists in U will be entries of X if and only if they have the same entries as Y. So, inside our pretense, we can say: yes, there is a list that lists exactly one list. We do so secure in the knowledge that, outside our pretense, this translates as a true statement about U. We got the translation to come out true by making certain assumptions about U. Different assumptions can yield a different result. For example, if there are lists in U not of type-1, but none of them are hereditarily invariant, then, no matter how much we might wish to broaden our perspective, H will just be T_1 and our proposition will translate as a falsehood. That will be alright, because we will be forced to deny the proposition within our pretext and that denial will translate as a true statement about our lists. Our pretense will still be a reliable guide to truth, but we will not have as many interesting things to say about lists.

Let us consider a simple universe of lists in which things turn out particularly well. Let $\emptyset_1$ and $\emptyset_2$ be blank lists (lists that list nothing). Make two lists that list only those blank lists: $\{\emptyset_1, \emptyset_2\}_1$ and $\{\emptyset_1, \emptyset_2\}_2$. Make two lists that list only those two lists: $\{\{\emptyset_1, \emptyset_2\}_1, \{\emptyset_1, \emptyset_2\}_2\}_1$ and $\{\{\emptyset_1, \emptyset_2\}_1, \{\emptyset_1, \emptyset_2\}_2\}_2$. We now have a little universe U consisting of two equivalent type-1 lists (the blank lists), two equivalent lists of type-1 lists, and two equivalent lists of lists of type-1 lists:

$$\begin{aligned} \emptyset_1 &\approx \emptyset_2 \\ \{\emptyset_1, \emptyset_2\}_1 &\approx \{\emptyset_1, \emptyset_2\}_2 \\ \{\{\emptyset_1, \emptyset_2\}_1, \{\emptyset_1, \emptyset_2\}_2\}_1 &\approx \{\{\emptyset_1, \emptyset_2\}_1, \{\emptyset_1, \emptyset_2\}_2\}_2. \end{aligned}$$

Since all our lists are hereditarily invariant, it is child's play to make-believe that $\approx$ is $=$. First, we limit our use of predicates that fail our invariance test. If $\{\emptyset_1, \emptyset_2\}_1$ is written in pencil, but $\{\emptyset_1, \emptyset_2\}_2$ is not, then the predicates

$$\_\_\_\_\_ \text{ is in pencil}$$

and

$$\_\_\_\_\_ \text{ is not in pencil}$$

are non-invariant. So it would spoil our pretense if we allowed ourselves to affirm or deny that our lists are in pencil. That would reveal that $\{\emptyset_1, \emptyset_2\}_1$ and $\{\emptyset_1, \emptyset_2\}_2$ are not really the same. We will insist instead that a list in U is just not the sort of thing that either is or is not in pencil. We will insist that a string of symbols such as

$$\{\emptyset_1, \emptyset_2\}_1 \text{ is in pencil } \vee \{\emptyset_1, \emptyset_2\}_1 \text{ is not in pencil}$$

is ungrammatical. Our next step is to pretend that all the numerical subscripts have disappeared and to announce that we have, not six lists, but three: $\emptyset$, $\{\emptyset\}$, and $\{\{\emptyset\}\}$. We can now elaborate on our pretense without fear of contradicting ourselves. Furthermore, we have a technique for translating make-believe results about our three lists into true statements about the six lists in U.

Note, for example, that in our make-believe world (the world inhabited by $\emptyset$, $\{\emptyset\}$, and $\{\{\emptyset\}\}$) no list has more than one entry. That is,

$$\forall X, Y(\exists Z(Z \text{ lists } X \wedge Z \text{ lists } Y) \rightarrow X = Y).$$

We interpret this as a true statement about our six lists:

$$\forall X, Y \in U(\exists Z \in U(Z \text{ lists } X \wedge Z \text{ lists } Y) \rightarrow X \approx Y).$$

We write 'U' rather than 'H' because H, the class of hereditarily invariant lists in U, just *is* U. The make-believe claim that no list has more than one entry translates as the true observation that U-lists appearing on the same U-list will have the same entries. Another example: in our make-believe world, lists that share an entry are the same. That is,

$$\forall X, Y(\exists Z(X \text{ lists } Z \wedge Y \text{ lists } Z) \rightarrow X = Y).$$

This translates as a true statement about U:

$$\forall X, Y \in U(\exists Z \in U(X \text{ lists } Z \wedge Y \text{ lists } Z) \rightarrow X \approx Y).$$

U-lists that share an entry share all their entries. Another example: in our make-believe world, lists that list the same lists are the same. That is,

$$\forall X, Y(\forall Z(X \text{ lists } Z \leftrightarrow Y \text{ lists } Z) \rightarrow X = Y).$$

This too we interpret as a true statement about U:

$$\forall X, Y \in U(\forall Z \in U(X \text{ lists } Z \leftrightarrow Y \text{ lists } Z) \rightarrow X \approx Y).$$

Since every entry of a U-list is a U-list, U-lists that list the same U-lists will, indeed, have the same entries. Thanks to our careful provisions, we can go on and on about our make-believe world

$$\emptyset, \{\emptyset\}, \{\{\emptyset\}\}$$

without any fear of contradicting ourselves. Even better, we have a systematic way to interpret each of our make-believe theorems as a true statement about the more complicated world of our six lists

$$\emptyset_1, \emptyset_2, \{\emptyset_1, \emptyset_2\}_1, \{\emptyset_1, \emptyset_2\}_2, \{\{\emptyset_1, \emptyset_2\}_1, \{\emptyset_1, \emptyset_2\}_2\}_1, \{\{\emptyset_1, \emptyset_2\}_1, \{\emptyset_1, \emptyset_2\}_2\}_2.$$

We have a simple and coherent pretense that is a reliable guide to some less simple truths.

6. Mirroring and contagion

Belief can influence pretense and pretense can influence belief. I might pretend that some proposition is true because I believe that some related proposition is true. When I make-believe that I have a dog, I might imagine that my dog brings me my slippers because I believe most dogs behave like that. I might imagine

My dog fetches my slippers

because I believe

Real dogs fetch slippers.

On the other hand, I might come to believe that some proposition is true because my pretense has led me to make-believe that some related proposition is true. I might come to believe that a dog will increase my happiness because I have run through the dog-ownership scenario in some detail and have been led to imagine myself as happier with a dog than I was without one. I might believe

I will be happier with a dog

because, in the course of a plausible make-believe, I came to imagine

My dog has made me happier.

The direction of flow

$$\text{Belief} \Longrightarrow \text{Pretense}$$

is known as MIRRORING. Make-believe scenarios will, under certain circumstances, reflect certain features of what we believe to be reality. The challenge is to specify what circumstances justify the reflection of what features and, furthermore, what counts as a reflection of what. The converse influence

$$\text{Pretense} \Longrightarrow \text{Belief}$$

has acquired a more pejorative label than it deserves: CONTAGION.[6] When a pretense *causes* me to modify my beliefs without providing a *reason* for doing so, this can, indeed, be a kind of pathology — rather like the transmission of disease. We, however, are interested in those cases where make-believe does provide reasons — and good reasons at that.[7] How does it do so? And what modifications of which beliefs can be justified in this way?

Logicians have a good understanding of both directions of flow (belief-to-pretense and pretense-to-belief) in cases like those we discussed above: cases where we pretend that equivalent objects are identical. The founding principles of any equivalence-is-identity pretense are:

Mirroring If you believe things are equivalent, then you should pretend they are identical.

Contagion If you find yourself pretending that things are identical, then you have a good reason to believe they are equivalent.

These are just instances of two more general principles. The idea is that we have an INTERPRETATION FUNCTION i that governs mirroring and contagion as follows:

Mirroring If you believe that $i(\phi)$ is true, then you should pretend that ϕ is true.

[6] For helpful discussions of mirroring, contagion, and, more generally, interactions between pretense and belief (as well as many references to a substantial literature), see [5], [6], [7], and [8].

[7] The role of imagination in the rational orientation of action was, by the way, one of John Dewey's favorite themes. "Only imaginative vision elicits the possibilities that are interwoven within the texture of the actual" [4, p. 345].

Contagion If you find yourself pretending that ϕ is true, then you have a good reason to believe that $i(\phi)$ is true.

If i translates equations as equivalences — that is, if $i(\alpha = \beta)$ is $\alpha \sim \beta$ where $\sim$ is the equivalence relation that is to impersonate identity — then we obtain our first two principles as special cases:

Mirroring If you believe that $\alpha \sim \beta$, then you should pretend that $\alpha = \beta$.

Contagion If you find yourself pretending that $\alpha = \beta$, then you have a good reason to believe that $\alpha \sim \beta$.

The challenge now is to say how i is to handle other formulas and to make sure that contagion is always benign: that is, to make sure that ϕ is derivable in our pretense only if $i(\phi)$ is derivable from propositions we actually believe. Happily for us, logicians have considerable experience rising to this very challenge. They are quite good at establishing INTERPRETABILITY.[8] We followed their example when we formulated the translation schemes we discussed above. We will now consider in more general terms how the logician's style of interpretation can help keep our pretenses coherent and can control contagion in ways that render it not just benign but illuminating.

7. Interpretability

In the discussion that follows, ϕ, ψ, χ are understood to be sentences while A, B, C are sets of sentences. $\vdash$ is a DERIVABILITY relation: $A \vdash \phi$ if and only if ϕ is derivable from sentences in A (or, to put it more briefly, A proves ϕ). We assume that $\vdash$ satisfies:

Transitivity If B proves every member of C and $C \vdash \phi$, then $B \vdash \phi$.

We say: A is INCONSISTENT if and only if A proves every sentence. Suppose i is a function that assigns sentences to sentences. $i[A]$ (the IMAGE

[8] Daniel Bonevac provides a clear introduction in [2, ch. 4]. If we view an ONTOLOGICAL REDUCTION as a justification of a particular sort of pretense (in which we make-believe there are things whose existence, in more serious moments, we deny), then we can read Bonevac's book as an exploration of how interpretation functions help us keep imagination coherent and contagion benign. For a taste of the logical literature, see [22].

of A under i) is $\{i(\phi) : \phi \in A\}$. i is an INTERPRETATION FUNCTION if and only if it satisfies the following two conditions.

Preservation of derivability If $A \vdash \phi$, then $i[A] \vdash i(\phi)$.
Preservation of inconsistency If A is inconsistent, so is $i[A]$.

If a conclusion is derivable from some premises, then the interpretation of the conclusion will be derivable from the interpretations of the premises. This is enough to guarantee that the interpretation of every sentence will be derivable from the interpretations of some inconsistent premises. That is, if A is inconsistent, then $i[A]$ will prove $i(\phi)$ for every sentence ϕ. It does not follow that $i[A]$ will prove every sentence.[9] So the second of our two conditions (preservation of inconsistency) does not follow from the first (preservation of derivability).

We say that A is INTERPRETABLE IN B if and only if, for some interpretation function i, B proves every member of $i[A]$. Here are two elementary theorems that will help to show why interpretability is of interest.

THEOREM 1. *If A is interpretable in B and i provides the interpretation, then A proves ϕ only if B proves $i(\phi)$.*

THEOREM 2. *If A is interpretable in B, then B is consistent only if A is consistent.*

Think of A as a set of pretenses while B is a set of beliefs. We really believe that the sentences in B are true, but only make-believe that the sentences in A are true. Suppose A is interpretable in B. Then Theorem 1 provides a recipe for a productive sort of contagion. If ϕ is derivable from our pretenses, then we have a good reason to make-believe that ϕ is true — but we also have a good reason to *believe* that $i(\phi)$ is true since we have shown that $i(\phi)$ is derivable from beliefs we already hold. Whenever we expand our stock of pretenses through inference, we identify sentences we ought to believe. On the other hand, Theorem 2 shows how our beliefs might regulate our pretenses. If you can translate your pretenses into beliefs, as we did above, then your pretenses will be coherent as long as your beliefs are.

[9] Suppose $\emptyset \vdash \top$. Let i interpret every sentence as $\top$. Then $i[A] \vdash i(\phi)$ no matter what A is. So i preserves derivability. Suppose $\{\top\}$ is consistent but $\{\bot\}$ is not. Then, since $i[\{\bot\}]$ is consistent, i does not preserve inconsistency.

To vindicate a pretense in this powerful way, we need a translation scheme that preserves both derivability and inconsistency. How would we arrange for a translation scheme to do that? As a first step in answering this question, we consider some conditions that are *necessary*. We determine that a condition is necessary by showing that it follows from our limited assumptions about interpretation functions and some modest assumptions about our logic. We assume that the operators $-$, $\wedge$, $\rightarrow$, and $\vee$ have the following properties.[10]

$(\phi \wedge \psi) \vdash \phi$.
$(\phi \wedge \psi) \vdash \psi$.
$\{\phi, \psi\} \vdash (\phi \wedge \psi)$.
$A \cup \{\phi\}$ is inconsistent if and only if $A \vdash -\phi$.
$A \cup \{\phi\} \vdash \psi$ if and only if $A \vdash (\phi \rightarrow \psi)$.
$(\phi \vee \psi) \vdash \chi$ if and only if $\phi \vdash \chi$ and $\psi \vdash \chi$.

From now on, we assume that i is an interpretation function. The following theorems show how i treats $-$, $\wedge$, $\rightarrow$, and $\vee$.

Theorem 3. $i(\phi \wedge \psi) \dashv\vdash (i(\phi) \wedge i(\psi))$.[11]

Theorem 4. $i(-\phi) \vdash -i(\phi)$.[12]

Theorem 5. $i(\phi \rightarrow \psi) \vdash (i(\phi) \rightarrow i(\psi))$.[13]

Theorem 6. $(i(\phi) \vee i(\psi)) \vdash i(\phi \vee \psi)$.[14]

If your translation scheme is to preserve both derivability and inconsistency, then (not too surprsingly) it must respect logical form to a substantial degree. Your translation of a conjunction must be deductively equivalent to the conjunction of your translations of the conjuncts. Your translation of the negation $-\phi$ must prove the negation of your

[10] We will sacrifice some brackets on the altar of readability. For example, in the statement of the first two properties, we write '$(\phi \wedge \psi)$' instead of '$\{(\phi \wedge \psi)\}$'. The latter would be more proper because $\vdash$ is supposed to be a relation between a *set* and a sentence. Note, by the way, that our operators are not necessarily classical. All the properties we attribute to them apply also to intuitionist connectives.

[11] $\theta \dashv\vdash \chi$ just means that θ and χ prove one another.

[12] The converse does not follow from our assumptions. Consider the modal logic S5. Let $i(\phi) = \Box\phi$. Then i preserves both derivability and inconsistency and, so, is an interpretation function in our sense. But $-i(\phi)$ (that is, $-\Box\phi$) does not prove $i(-\phi)$ (that is, $\Box - \phi$).

[13] Again, the converse is unprovable: $\Box\phi \rightarrow \Box\psi$ does not prove $\Box(\phi \rightarrow \psi)$ in S5.

[14] Yet again, the converse is unprovable: $\Box(\phi \vee \psi)$ does not prove $\Box\phi \vee \Box\psi$ in S5.

translation of ϕ. And so on. The general point is that if f behaves like the logical operators we have considered, then we will need derivability to hold in at least one direction between $i(f(\phi_1, \ldots, \phi_n))$ and $f(i(\phi_1), \ldots, i(\phi_n))$. We will need either

$$i(f(\phi_1, \ldots, \phi_n)) \vdash f(i(\phi_1), \ldots, i(\phi_n))$$

or

$$f(i(\phi_1), \ldots, i(\phi_n)) \vdash i(f(\phi_1, \ldots, \phi_n))$$

if not both. One effective way to guarantee inter-derivability is to insist on *identity*:

$$i(f(\phi_1, \ldots, \phi_n)) = f(i(\phi_1), \ldots, i(\phi_n)).$$

That is:

$$\begin{aligned} i(-\phi) &= -i(\phi) \\ i(\phi \wedge \psi) &= (i(\phi) \wedge i(\psi)) \\ i(\phi \rightarrow \psi) &= (i(\phi) \rightarrow i(\psi)) \\ i(\phi \vee \psi) &= (i(\phi) \vee i(\psi)). \end{aligned}$$

This is one option. In what follows, we do not assume that we have chosen this option.

As we saw above, we may want our translation scheme to show less respect for two other components of logical form: the identity relation and quantifiers. It may be useful to replace equations with equivalences and unbounded quantifiers with bounded ones:

$$\begin{aligned} i(\alpha = \beta) &= \alpha \sim \beta \\ i(\forall \alpha\ \phi) &= \forall \alpha \in M\ i(\phi) \\ i(\exists \alpha\ \phi) &= \exists \alpha \in M\ i(\phi). \end{aligned}$$

This treatment of quantifiers will be safe as long as we can confirm that M is not empty. (If M were empty, every existential generalization would translate as a falsehood.) As we have already seen, it may not be quite that easy to guarantee that our interpretation of $=$ as $\sim$ is safe.

It will, presumably, be a theorem of our logic that identity is an equivalence relation (reflexive, symmetric, and transitive). So, if i preserves derivability, we will be able to prove that $\sim$ is an equivalence relation in M [16, p. 60]. Assume, for example, that the symmetry of identity is a theorem:

$$\forall x, y(x = y \rightarrow y = x).$$

Then, since i preserves derivability:

$$\forall x, y \in M\ i(x = y \rightarrow y = x).$$

By Theorem 5,

$$\forall x, y \in M\ i(x = y \rightarrow y = x) \vdash \forall x, y \in M(x \sim y \rightarrow y \sim x).$$

So

$$\vdash \forall x, y \in M(x \sim y \rightarrow y \sim x).$$

That is, we can prove that $\sim$ is symmetric in M. We can do the same for reflexivity and transitivity. The point is: a necessary condition for i to preserve derivability is that $\sim$ be an equivalence relation in M.

Another necessary condition is that each instance of Leibniz's Law have a provable translation — which, as we shall now see, places further demands on our translation scheme. Assume:

$$\forall x, y((\phi(x) \wedge x = y) \rightarrow \phi(y)).$$

Then, by Theorems 3 and 5 and the preservation of derivability:

$$\forall x, y \in M((i(\phi(x)) \wedge x \sim y) \rightarrow i(\phi(y))).$$

This means: for each formula $\phi(\alpha)$, the translation $i(\phi(\alpha))$ must be invariant in M [16, p. 61]. We saw above how we might arrange for this through a careful choice of M and special restrictions on what counts as a formula.

To sum up: we have reviewed some conditions that must be satisfied if our translation scheme is to preserve derivability and inconsistency. Our scheme must respect logical form at least to the extent of satisfying Theorems 3–6. Our scheme can show less respect for quantifiers and identity: it can place a bound M on our quantifiers and can interpret '=' as some relation $\sim$. But then M must be non-empty, $\sim$ must be an equivalence relation in M, and any formulas we count as grammatical must have a translation that is invariant in M.

8. Conclusion

Formal logic supplies tools for the rational guidance of belief by imagination. These tools may help us better understand the pretense-belief connection both in and out of the formal sciences.

References

[1] Angelelli, I., "Adventures of abstraction", *Poznan Studies in the Philosophy of the Sciences and the Humanities* 82 (2004): 11–35.

[2] Bonevac, D. A., *Reduction in the Abstract Sciences*, Hackett, Indianapolis, 1982.

[3] Burgess, J. P., *Fixing Frege*, Princeton University Press, Princeton, New Jersey, 2005.

[4] Dewey, J., *Art as Experience*, Minton, Balch and Company, New York, 1934.

[5] Gendler, T. S., "On the relation between pretense and belief", pages 125–141 in D. M. Lopes and M. Kieran (eds.), *Imagination, Philosophy and the Arts*, Routledge, London, 2003. DOI: 10.1093/acprof:oso/9780199589760.003.0008

[6] Gendler, T. S., "Imaginative contagion", *Metaphilosophy* 37, 2 (2006): 183–203. DOI: 10.1111/j.1467-9973.2006.00430.x

[7] Gendler, T. S., "Imagination", In E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*, plato.stanford.edu, 2013.

[8] Leeuwen, N. V., "The meanings of "imagine", part II: Attitude and action", *Philosophy Compass* 9, 11 (2014): 791–802. DOI: 10.1111/phc3.12141

[9] Lorenzen, P., "Equality and abstraction", *Ratio* 4 (1962): 85–90.

[10] Lorenzen, P., *Formal Logic*, D. Reidel, Dordrecht-Holland, 1965.

[11] Mancosu, P., "Grundlagen, Section 64: Frege's discussion of definitions by abstraction in historical context", *History and Philosophy of Logic* 36, 1 (2015): 62–89. DOI: 10.1080/01445340.2014.967950

[12] Pasch, M., "Die Begründung der Mathematik und die implizite Definition: Ein Zusammenhang mit der Lehre vom Als-Ob", *Annalen der Philosophie* 2 (1921): 145–162.

[13] Pasch, M., "Implicit definition and the proper grounding of mathematics", pages 95–107 in S. Pollard (ed.), *Essays on the Foundations of Mathematics by Moritz Pasch*, chapter 4, Springer, New York, 2010. DOI: 10.1007/978-90-481-9416-2_5

[14] Pollard, S., "What is abstraction?", *Noûs* 21, 2 (1987): 233–240. DOI: 10.2307/2214916

[15] Pollard, S., "Weyl on sets and abstraction", *Philosophical Studies* 53 (1988): 131–140. DOI: 10.1007/bf00355680

[16] Pollard, S., *Philosophical Introduction to Set Theory*, University of Notre Dame Press, Notre Dame, Indiana, 1990.

[17] Pollard, S., "'As if' reasoning in Vaihinger and Pasch", *Erkenntnis* 73 (2010): 83–95. DOI: 10.1007/s10670-009-9205-7

[18] Pollard, S., *A Mathematical Prelude to the Philosophy of Mathematics*, Springer, New York, 2014. DOI: 10.1007/978-3-319-05816-0

[19] Quine, W. V. O., *From a Logical Point of View*, Harvard University Press, Cambridge, Massachusetts, 1980.

[20] Scott, D., "More on the axiom of extensionality", pages 115–131 in Y. Bar-Hillel, E. I. J. Poznanski, A. Robinson, and M. O. Rabin (eds.), *Essays on the Foundations of Mathematics*, Hebrew University, Jerusalem, 1966.

[21] Simons, P., "What is abstraction and what is it good for?", pages 17–40 in A. D. Irvine (ed.), *Physicalism in Mathematics*, Kluwer, Dordrecht, 1990. DOI: 10.1007/978-94-009-1902-0_2

[22] Visser, A., "An overview of interpretability logic", pages 307–359 in M. Kracht, M. De Rijke, and M. Zakhary (eds.), *Advances in Modal Logic*, volume 1, CSLI Publications, Stanford, CA, 1998.

[23] Weyl, H., *Philosophy of Mathematics and Natural Science*, Princeton University Press, Princeton, New Jersey, 1949.

Stephen Pollard
Department of Philosophy and Religion
Truman State University
Kirksville, Missouri 63501 USA
spollard@truman.edu