Substitution and Subtraction

DHOEK@PRINCETON.EDU, WWW.DANIELHOEK.COM, 11 DECEMBER 2019

> "Identity is a rather puzzling thing at first sight. When you say "Scott is the author of Waverly", you are half-tempted to think there are two people, one of whom is Scott and the other the author of Waverly, and they happen to be the same. That is obviously absurd, but that is the sort of way one is always tempted to deal with identity."
>
> — Bertrand Russell, *The Philosophy of Logical Atomism*

> "When we quote a man's utterance directly we report it almost as we might a bird call… On the other hand in indirect quotation we project ourselves into what, from their remarks and other indications, we imagine the speaker's state of mind to have been, and then we say what, in our language, is natural and relevant for us in the state thus feigned."
>
> — W.V. Quine, *Word and Object*

Saul on Substitution

We are all familiar with substitution failures of the following kind:

1 a) Lois believes Superman can fly.

b) Lois believes Clark can fly.

In the Superman story, (1a) is intuitively false but (1b) is true, even though the only difference is the substitution of a coreferential term. What's worse, the received, Millian view of proper names dictates that "Superman" and "Clark" do not only corefer but are actually *synonymous*, which means that basic considerations of compositionally would seem to show that (1a) and (1b) are also synonymous.

Now ever since Frege originally brought up this problem, it has been customary to say that the (apparent) difference between (1a) and (1b) is somehow connected to the fact that the names occur in a special, *opaque* context, generated in this case by the attitude verb "believe". In regular, transparent contexts these apparent failures of substitution are supposed to be impossible. For example, (2a) and (2b) seem to have identical truth conditions:

2 a) Superman is from another planet.

b) Clark Kent is from another planet.

In her short paper "Substitution and Simple Sentences," Jeniffer Saul (1997) challenges that orthodoxy, offering a long list of examples in which coreferential substitutions in a supposedly transparent context intuitively make for a very clear difference in meaning (see also Bezuidenhout 1996):

3 a) Clark Kent went into the phone booth, and Superman came out.

b) Clark Kent went into the phone booth, and Clark Kent came out.

4 a) She made a date with Superman, but found herself having dinner with Clark Kent.
b) She made a date with Superman, but found herself having dinner with Superman.
5 a) Clark Kent always arrived at the scene just after one of Superman's daring rescues.
b) Superman always arrived at the scene just after one of Clark Kent's daring rescues.
6 a) He kicked Clark Kent once, but he never kicked Superman.
b) He kicked Superman once, but he never kicked Clark Kent.
7 a) I never made it to Leningrad, but I visited St. Petersburg last week.
b) I never made it to Leningrad, but I visited Leningrad last week.

Saul argues there is a continuity between examples like (3-7) and example (1), and that a satisfactory explanation of the data should provide a uniform treatment of all these cases.

But what might such a treatment look like? Saul persuasively argues that a semantic explanation for the contrasts (3-7) is particularly unlikely to be forthcoming. So in the paper, she tentatively takes these examples to support the view that we need a pragmatic treatment for all these contrasts. She notes that this in line with Nathan Salmon's position that the perceived contrast between (1a) and (1b) is entirely pragmatic. Or to put Saul's thought here a different we way: whatever happens, it looks like we are going to need a pragmatic story to handle (3-7). It seems likely that story will cover (1) too, so no need to tie ourselves into special semantic knots over attitude reports.

But as Saul later recognised in her book on this topic, a purely pragmatic approach also faces apparently insuperable difficulties. Most Gricean derivations takes the following form:

Premise 1: Taken literally, the speaker said that *p*.
Premise 2: [Background knowledge about the speaker and the conversational aims]
Conclusion. The speaker must have meant that *q*.

But on the assumption that, for instance, (3a) and (3b) mean exactly the same thing, the inputs of this derivation are identical in both cases. And that implies that any rational conclusion that can be drawn when the speaker utters (3a) can also be drawn if they utter (3b) instead. So it is *a priori* impossible to give a systematic pragmatic story of this kind about how (3a) and (3b) are supposed to come apart.

(One possibility is that these might be *manner* implicatures. Manner implicatures are supposed to arise from the particular way the speaker puts things, and are therefore the only kinds of conversational implicatures within Grice's framework that are not "detachable" (an implicature is *detachable* if it would still be there if the speaker had used a different, synonymous form of expression). But it is not clear how that idea is to be fleshed out.)

What If We're Only Pretending There's a Difference in Reference?

Now for those who want to take Jeniffer Saul's lesson on board, and provide a uniform treatment for the contrast in (1-7), on the face of it the truly vast literature on Frege's Puzzle has surprisingly little to offer. The distinction between opaque and transparent contexts is so deeply etched into the philosophical consciousness that the majority of explanations of the contrast in (1) appeal turn on properties of the verb "believes" in particular or of attitude reports in general. For this reason, it is difficult to see how those explanations could extend to account for (3-7).

A notable exception is Mark Crimmins' (1998) proposal to understand substitution failures as due to a specific kind of prop-oriented make-believe. On Crimmins' diagnosis, problems about substitution arise whenever we appeal to a make-believe game in which it is pretended that two things are different that are in actual fact identical.

To warm people up to that view, Crimmins points out that there are other plausible cases of make-believe in which there are more people involved in the make-believe than we are really talking about:

8) Ann is as clever as Holmes and more modest than Watson.

9) Elijah believes that Santa is overworked.

In (8) we're make-believedly referring to three people, but only talking about Ann. In (9) we're make-believedly referring to two people but only talking about Elijah and the contents of his beliefs. Could it be that (1a) is just another example where we are make-believedly talking about three people Lois, Clark and Superman, even though really there are only two? We could add Donnellan's Martini cases to the list of phenomena that could potentially be explained in terms of make-believe and that seem to bear some relation to Frege's Puzzle:

10) [*Jill is not in the room*]. Jill thinks we should hire the man with the martini over there.

Here we can get a reading *Jill thinks we should hire that guy* by subtracting the presupposition *That guy is the man with the martini over there*, with the subject matter *Who does Jill think we should hire*.

Now here is roughly how this idea could be fleshed out in the case of (1a):

p: *Lois believes Superman can fly*

q: *Superman and Clark Kent are different people. Superman is the damsel-saving cape-wearing hero of Metropolis, and Clark Kent is the pencil-sharpening tie-toting dimwit of the Daily Planet.*

S: *What does Lois's think about the damsel-saving cape-wearing hero of Metropolis?*

r: *Lois believes that the the damsel-saving cape-wearing hero of Metropolis can fly.*

To argue that $\mathrm{S}(p \upharpoonright q) = r$, we need to check the usual three criteria:

i) *Aboutness*: r is wholly about S; this seems unproblematic in this case.

ii) *Equivalence*: $p \upharpoonright q = r \upharpoonright q$; no issue here either. In fact, suppose we formulated p, q and r in first-order logic in the natural way as ϕ, ψ and χ respectively, then $\psi \supset (\phi \equiv \chi)$ will be a logical truth.

iii) *Independence*: r has no bearing on S; there's the rub: r seems like a necessary falsehood, which would mean $\mathrm{S}(p \upharpoonright q)$ ends up being a partial proposition without a truth-value at any world.

With regard to (ii), note that we get a bit of a contrast here between (1a) and (1b). For if we formalise the sentence (1b) in the obvious way as ϕ', then $\psi \supset (\phi' \equiv \chi)$ does *not* come out as a logical truth. Now it is not clear if and how this is going to help: after all, the exculpature account is formulated at the level of propositions, not sentences. But it should give us some hope that, if we somehow find a way to think of q contingent, that may give us the key to accounting for these contrasts.

Ignoring those problems for now, we do indeed get an analogous treatment of (5a):

p': *Clark Kent went into the phone booth, and Superman came out.*

q: *Superman and Clark Kent are different people. Superman is the damsel-saving cape-wearing hero of Metropolis, and Clark Kent is the pencil-sharpening tie-toting dimwit of the Daily Planet.*

S′: *What did the scene at the phone booth look like?*

r': *A damsel-saving capewearing hero entered the phonebooth, and a pencil-sharpening tie-toting dimwit exited the phone booth.*

The comments about (1a) carry over pretty much verbatim here. (For the equivalence to go through, we do need to lean on the uniqueness assumptions from the definite descriptions: there's only one (relevant) damsel-saving capewearing hero, and only one pencil-sharpening tie-toting dimwit.)

Contingent Identities

There are different approaches one might take to the problem of independence:

- *Contingentism about Identities* (Bacon and Russell 2017). This view denies it is metaphysically necessary that *Hesperus is Phosphorus*, while still holding on to a version of Leibniz' Law.
- *Variabilism* (Cumming 2008). On this approach, propositions are world-assignment pairs rather than sets of worlds, where the assignments assign names to individuals. The presupposition q will be a set of pairs that do not make "Superman" and "Clark" corefer.
- *Diagonalisation* (Stalnaker 1978). On this approach, a name α basically acquires the meaning 'the referent of α' for pragmatic reasons.

The first two of these strategies lead to a mixed semantic-pragmatic explanation. The third appeals to pragmatic principles that go beyond the usual Gricean arsenal.

Humberstone on Subject Matter Subtraction

Recall where we left things with the hyperintensional strategy: the exculpature view of subtraction basically reduces the problem of finding the remainder $P^S - Q^T$ to the problem of finding the *subject matter* U of that remainder R^U. The most natural hypothesis is that U = S – T, which would make an account of subject matter subtraction the key to a hyperintensional theory of propositional subtraction. However, on the view of subject matters as partitions, it is not straightforward to get such a theory. In particular, we saw that even where circumstances are propitious, there is typically more than one subject matter U such that (i) U is orthogonal to T, and also (ii) some subject matter UT = S.

That is where Humberstone's account comes in. He proposes a return to a more restrictive, less promiscuous notion of subject matter, which prunes away at the unwanted abundance of possible remainders. Essentially, the idea here is to go with a restrictive version of David Lewis's initial analysis of subject matter: let subject matters be intensionally individuated, concrete parts of the world, rather than allowing arbitrary partitions. The cost of this approach is generality. The rewards are mathematical elegance, and a new theory of subtraction.

Parts of the World

To get the Humberstonian theory of subject matter off the ground, we need to help ourselves to the mereology of the world: think of the universe as a very large object, and consider its parts. As Humberstone emphasises, there are various different ways in which one might conceive of this starting point. If one thinks of the world as a very large place, its parts correspond to spatial regions. If one thinks of the world as a very long story, its parts are temporal episodes. If one thinks of the world as the totality of things, its parts may be collections of things.[1]

Correspondingly, there are different kinds of propositional parts. Consider the proposition

> *P*: *Olga, Masha and Igor only sleep during the day*

The spatial parts of *P* correspond to spatial regions of the world:

> $P_{\text{the garden}}$: *Olga, Masha and Igor only sleep in the garden during the day*
>
> P_{Odessa}: *When in Odessa, Olga, Masha and Igor only sleep during the day*
>
> P_{Kiev^*}: *When in Kiev, Olga, Masha and Igor only sleep during the day*

[1] For this latter, 'thing-based' or 'reic' version of the account to have a chance at being compatible with the *Recombination* principle imposed below, we will need the collections to contain all the things that are part of the things they contain, and we also need it to be the case that things have their parts essentially. That is a bit much, but it's not the last bold simplifying assumption we'll need to make here. A way to skirt the issue would be to think of the world as a collection of partless atoms instead, and of world-parts as subsets of that collection.

P's temporal parts correspond to moments or stretches of time:

P_{2003}: *In 2003, Olga, Masha and Igor only slept during the day*

P_{tonight} : *Olga, Masha and Igor will not sleep tonight*

$P_{\text{2-4am}}$: *Olga, Masha and Igor never sleep between 2am and 4am*

$P_{\text{2-4pm}}$: $\top$

And P's reic parts correspond to sets of objects:

P_{Masha}: *Masha only sleeps during the day*

P_{women}: *Olga and Masha only sleep during the day*

P_{Dmitri}: $\top$

World Parts

Instead of thinking of possible worlds as points or unstructured entities, worlds in this formalism will be really big objects with parts. Depending on the intended application, these may be spatial, temporal or reic parts. Each world w has a space M_w of ***extensional world parts***, ordered by the parthood relation $\sqsubseteq_w$. We assume $\sqsubseteq_w$ obeys classical mereology. That is, the parts form a complete Boolean Algebra equipped with fusion, intersection and complementation operators $\sqcup_w$, $\sqcap_w$ and $\circledast_w$. The top part of a world w is the part containing all the parts of w — which is to say, w itself. Its bottom part is always the empty part $\square$, which we will take to be identical for every possible worlds.

In addition, there is a space M of ***intensional world parts***. Depending on whether we are doing spatial, temporal or reic parts, think of the members of M as world-neutral regions of space, time intervals, or collections of things. These intensional parts get filled in differently, by different extensional parts at different worlds. Formally then, the members m of M will be functions that take a world w to the extensional part of w that fills in m at w. For instance, if the function n represents the region *New Jersey*, then for any world w, $n(w)$ is the content of that region at w: everything that, at w, is happening in New Jersey. Every part of every world is picked out by some $m \in M$. If a part m exists only contingently, we'll have that $m(w) = \square$ for some worlds w but not others.

Humberstone then defines the intensional parthood relation $\sqsubseteq$ as follows:

$m \sqsubseteq n$ if and only if for all worlds w, $m(w) \sqsubseteq_w n(w)$

This automatically gives us fusion, intersection and complementation operators $\sqcup$, $\sqcap$ and $\circledast$ on M, as well as a top part $\blacksquare$ and an empty part $\square$. (For every world w, $\blacksquare(w) = w$, while $\square(w) = \square$). Two parts m, n are *disjoint* iff $m \sqcap n = \square$. We can also easily define a ***subtraction*** of intensional parts:

$m - n = m \sqcap n^* =$ the largest part of m that is disjoint with n

Now we will think of the members of M as functions from possible worlds w to extensional parts of each world, with the feature that every extensional part of a world is the value of $m(w)$ for some $m \in M$. We can recover the extensional mereology of a particular world from M. The set $M(w)$ of ***extensional world parts*** of w is just

$M(w) = \{ m(w) : m \in M \}$

And the parthood relation $\sqsubseteq_w$ can also be characterised in terms of $\sqsubseteq$:

for any $x, y \in M(w)$, $x \sqsubseteq_w y$ if and only if, for some $m, n \in M$, $m \sqsubseteq n$, $m(w) = x$ and $n(w) = y$

Likewise $\sqcup_w$, $\sqcap_w$ and $*_w$ can be defined in terms of $\sqsubseteq$.

From Parts to Partitions

Sometimes two worlds *agree* with respect to a particular intensional part. For instance two worlds w and v might agree on everything that happened yesterday. If m is the intensional part representing *yesterday*, we write this $m(w) \approx m(v)$. How to interpret '≈' may depend on your metaphysical proclivities. On one way of thinking about it, the possible worlds are different ways of combining the exact same parts. If you have three handles and seven blades, there are 21 possible knives out of exactly those handles and blades. Likewise if you take five possible Mondays, five possible Tuesdays, and five possible versions of every day in history, you can combine them into 5^N possible worlds, where N is the total number of days. On this way of viewing the matter, ≈ is numerical identity.

Alternatively, in a more Lewisian spirit, we can think of $m(w) \approx m(v)$ as saying that $m(w)$ and $m(v)$ are *intrinsic duplicates*. Either way, all members $m \in M$ induce a partition on the set of possible worlds:

$\mathrm{S}[m] = \{ \{w : m(w) \approx m(v) \} : v \in \Omega \}$

Humberstone uses the notation $\sigma(m)$). Now it is generally accepted that parts satisfy the principle of ***upwards difference transmission***: you cannot make a change to a part without changing the whole it belongs to. It follows from this that whenever $m \sqsubseteq n$, it is also true that $\mathrm{S}[m]$ is part of $\mathrm{S}[n]$.

Humberstone makes two further assumptions that can be stated in subject matter terms. The first assumption is that the state of the fusion $m \sqcup n$ is determined by the state of m and of n:

> **Locality**: For any parts m and n, if $m(w) \approx m(v)$ and $n(w) \approx n(v)$ then $m \sqcup n(w) \approx m \sqcup n(v)$.
> Or: $\mathrm{S}[m \sqcup n]$ is part of $\mathrm{S}[m] \wedge \mathrm{S}[n]$.

$\mathrm{S}[m]$ and $\mathrm{S}[n]$ are both parts of $\mathrm{S}[m \sqcup n]$ because m and n are part of $m \sqcup n$. So it follows from *Locality* that $\mathrm{S}[m \sqcup n] = \mathrm{S}[m] \wedge \mathrm{S}[n]$. The second assumption, based on a Lewisian principle of plenitude, says that any two states of disjoint parts are consistent with one another:

> **Recombination**: If m and n are disjoint, then for any w and v, there exists a world u such that $m(w) \approx m(u)$ and $n(u) \approx n(v)$. Or: for disjoint m and n, $\mathsf{S}[m]$ is orthogonal to $\mathsf{S}[n]$.

From *Recombination*, we get that $\mathsf{S}[m \sqcap n] = \mathsf{S}[m] \vee \mathsf{S}[n]$, where $\mathsf{S}[m] \vee \mathsf{S}[n]$ is understood as the overlap (greatest common coarsening) of $\mathsf{S}[m]$ and $\mathsf{S}[n]$. (This can be established as follows. Since $m \sqcap n$ is part of both m and n, $\mathsf{S}[m \sqcap n]$ must be part of both $\mathsf{S}[m]$ and $\mathsf{S}[n]$. To show that $\mathsf{S}[m \sqcap n]$ is the greatest common part we show that, if T is any partition intermediate between $\mathsf{S}[m \sqcap n]$ and $\mathsf{S}[m]$, then T cannot be part of $\mathsf{S}[n]$. Note that by *Recombination* $\mathsf{S}[m]$, $\mathsf{S}[m \sqcap n]$ and T must all be orthogonal to the subject matter $\mathsf{S}[m - n]$. Moreover $\mathsf{S}[m] = \mathsf{S}[m - n] \wedge \mathsf{S}[m \sqcap n]$ by our previous result. But since T is strictly more fine-grained than $\mathsf{S}[m \sqcap n]$ and still orthogonal to $\mathsf{S}[m - n]$, it follows that $\mathsf{S}[m - n] \wedge \mathsf{T}$ is not equal to $\mathsf{S}[m]$, being too fine-grained. Hence T is not part of $\mathsf{S}[n]$.)

Putting it all together, these two assumptions thus yield a perfect correspondence between the lattice of partition subject matters and the lattice of intensional world parts. In particular, that means we now have the final missing puzzle piece to complete the hyperintensional analysis of logical subtraction, namely a natural way to subtract subject matters from one another:

$$\mathsf{S}[m] - \mathsf{S}[n] := \mathsf{S}[m - n]$$

From World Mereology to Propositional Mereology

Now that we have the Humberstonian notion of subject matter on the table, we can use it to divide proportions up into parts as well. *Propositions*, here, will just be sets of worlds:

> The ***part of*** p ***about*** m, written p_m, is the proposition $\{ w : m(w) \approx m(v)$ for some $v \in p \}$.
>
> A proposition p is ***wholly about*** m just in case $p_m = p$. The ***habitat*** of a proposition p, written $hab(p)$, is the $\sqsubseteq$-smallest world part m such that $p_m = p$.

Thus Humberstone's formalism gives us a purely *intensional* notion of subject matter: the subject matter of any proposition is the smallest intensional world part needed to ascertain the truth of the proposition. (Or perhaps more accurately, Humberstone gives us a family of intensional notions of subject matter: there is a propositions spatial habitat, its temporal habitat and its reic habitat.)

Almost everything we need is now in place to define a notion of propositional subtraction. Not all of a proposition's parts is suitable for subtraction. For instance, the proposition *The weather in Chicago is the same as in New York* does not neatly divide into a part about Chicago and a part about New York.

(Applying the definitions above, the parts about both cities will just be tautologies). To address this, we restrict attention to parts which are extricable:

> A part p_m of p is ***extricable*** if and only if $p = (p_m \wedge p_{m^*})$

This immediately suggests the following simple definition of subtraction:

> If p_m is any extricable part of p, the ***remainder*** of p after subtracting p_m is $p - p_m =_{\text{df}} p_{m^*}$

This account vindicates all of Jaeger's basic desiderata. Both the subtracted proposition and the remainder are parts of the whole. The conjunction of the remainder and the subtracted proposition gives us back the initial proposition. And the subtracted part p_m and the remainder p_{m^*} are cleanly separate: they do not share any parts, so that nothing of the subtracted part "is left" in the remainder. That is because p_m and p_{m^*} must always have disjoint habitats. The fusion of those habitats is equal to the habitat of p.

It can be shown that the extricable parts of any given proposition form a Boolean algebra. The definition is fairly straightforwardly extended to the case where the subtracted proposition is not entailed by the whole:

> If q is any extricable part of $(p \,\&\, q)$, then the ***generalised remainder*** of p after subtracting q
> is $p \parr q =_{\text{df}} (p \,\&\, q)_{hab(q)^*}$

Some Remaining Remarks:

- For any world parts m and n, and any proposition p, $p_{m \sqcap n} = (p_m)_n = (p_n)_m$.
- The extricable parts of any proposition p form a Boolean Algebra under the parthood order.
- If q is an extricable part of $(p \,\&\, q)$, r is an extricable part of $(p \,\&\, r)$, and both are extricable parts of $(p \,\&\, q \,\&\, r)$, then $(p \parr q) \parr r = (p \parr r) \parr q = p \parr (q \,\&\, r)$
- Applications of this account of subtraction of numbers, fictional objects, and the subjects of paintings seems feasible. But most metaphors and loose talk appear to be out of reach.
- What about that Wittgenstein example of subtracting your arm going up from raising your arm?