# Substitutional Quantifiers

John MacFarlane—Philosophy 142

April 1, 2008

## 1 Use and mention

So far, we've been a bit casual about the distinction between use and mention. Consider the word 'sentence'. In this sentence, I am *using* it. But when I talk about the word 'sentence', I am *mentioning* it without using it.

It is customary to use quotation marks to indicate that one is mentioning a word. Sometimes italics are used to the same effect. And sometimes the author just relies on the reader to figure out what is being used and what is being mentioned. Usually this last approach works fine, but we are getting to issues where the use-mention distinction is very important, and failures to mark it explicitly can lead to fallacious reasoning. So we will start being more careful about it. To refresh your memory, then:

- 'net' is part of 'clarinet', but a net is not part of a clarinet.
- Boston is a city. 'Boston' is the name of a city. "Boston" is the name of a name of a city.
- An hour is longer than a minute, but 'minute' is longer than 'hour'.

In giving semantic clauses for logical connectives and operators, we have used phrasing like this:

> Where $\phi$ and $\psi$ are formulas, $\phi \wedge \psi$ is true in a model $\mathcal{M}$ iff $\phi$ is true in $\mathcal{M}$ and $\psi$ is true in $\mathcal{M}$.

How can we rephrase this in a way that is more careful about use and mention? Well, we might try:

> Where $\phi$ and $\psi$ are formulas, '$\phi \wedge \psi$' is true in a model $\mathcal{M}$ iff '$\phi$' is true in $\mathcal{M}$ and '$\psi$' is true in $\mathcal{M}$.

But this won't work! Remember, $\phi$ and $\psi$ are variables whose *values* are formulas. They are not formulas themselves. The expression

(1) '$\phi \wedge \psi$'

denotes the sequence of symbols:

(2) $\phi \wedge \psi$

which is not, itself, a formula.

How can we say what we want to say, then? Well, we could say this:

> Where $\phi$ and $\psi$ are formulas, the formula consisting of $\phi$ concatenated with '$\wedge$' concatenated with $\psi$ is true in a model $\mathcal{M}$ iff $\phi$ is true in $\mathcal{M}$ and $\psi$ is true in $\mathcal{M}$.

We could make this simpler by introducing a notation for concatenation:

‘$x \frown y$’ denotes $x$ concatenated with $y$.

But the result is still pretty ugly:

Where $\phi$ and $\psi$ are formulas, $\phi \frown$ ‘ $\wedge$ ’ $\frown \psi$ is true in a model $\mathcal{M}$ iff $\phi$ is true in $\mathcal{M}$ and $\psi$ is true in $\mathcal{M}$.

For this reason, Quine invented a device of “quasiquotation” or “corner quotes.” Using corner quotes, we could write our semantic clause like this:

Where $\phi$ and $\psi$ are formulas, $\ulcorner \phi \wedge \psi \urcorner$ is true in a model $\mathcal{M}$ iff $\phi$ is true in $\mathcal{M}$ and $\psi$ is true in $\mathcal{M}$.

You can think of corner quotes as a notational shortcut:

(3) $\ulcorner \phi \wedge \psi \urcorner$

*means* just the same as

(4) $\phi \frown$ ‘ $\wedge$ ’ $\frown \psi$

More intuitively, you can understand the corner quote notation as follows. A corner-quote expression always denotes another expression, relative to an assignment of values to its variables. To find out what expression it denotes on a given assignment of values to these variables, first replace the variables inside corner quotes with their values (which should be expressions), then convert the corner quotes to regular quotes. So, for example, when $\phi$ = ‘Cats are furry’ and $\psi$ = ‘Snow is black’, $\ulcorner (\phi \wedge \psi) \urcorner$ = ‘(Cats are furry $\wedge$ Snow is black)’.

Suppose the baby is learning to talk, and says, ‘I like blue,’ ‘I like red,’ ‘I like white,’ ‘I like green,’ and so on for all the color words he knows. You want to report what happened in a compact way. You want to say something like this:

(5) For every color word $C$, he said, ‘I like $C$.’

Strictly speaking, though, this reports him as having said, ‘I like $C$’, not ‘I like blue’, etc. To get the desired reading, you can use corner quotes:

(6) For every color word $C$, he said, $\ulcorner$I like $C\urcorner$.

*Exercises:*

1.1 Rewrite the following using corner-quote notation:

(a) $\phi \frown$ '$+$' $\frown \psi \frown$ '$=$' $\frown \psi$

(b) '$\forall$' $\frown \alpha \frown \phi$

1.2 Rewrite the following using regular quotes and the concatenation sign:

(a) $\ulcorner \exists \alpha(\phi \wedge \psi) \urcorner$

(b) $\ulcorner \phi \supset \phi \urcorner$

(c) $\ulcorner \phi \urcorner$

1.3 Write the expression denoted by the following terms, under an assignment of '$Fx$' to $\phi$, '$(Fx \supset Gx)$' to $\psi$, and '$x$' to $\alpha$:

(a) $\ulcorner \phi \wedge \psi \urcorner$

(b) $\ulcorner \forall \alpha(\psi \supset \phi) \urcorner$

## 2 Substitutional quantifiers

In giving the semantics of regular (*objectual*) quantifiers, we say which assignments of objects from the domain to the variables must make the embedded open formula true in order for the whole quantified formula to be true. We say, in effect: this formula (with a quantifier binding the variable $\alpha$) is true just in case this open formula (with the $\alpha$ free) is *true on* all/some assignments of values to $\alpha$.

*Substitutional* quantifiers have a different semantics. We say, in effect: this formula (with a quantifier binding $\alpha$) is true just in case all/some of the formulas you'd get by replacing $\alpha$ with a *name* are true.

Note that the semantics for substitutional quantifiers do not require us to talk about variable assignments at all. If we *only* had substitutional quantifiers in our language, and no objectual quantifiers, we could do without assignments. But it will often be useful to consider both kinds of quantifiers at the same time, so I'll explain the semantics of substitutional quantifiers in the framework of models and assignments we've been using all along. For the same reason, we'll use different symbols for the substitutional quantifiers: '$\Sigma$' for the existential and '$\Pi$' for the universal.

**Substitutional quantifiers**

$\ulcorner \Sigma\alpha\phi \urcorner$ is true in $\mathcal{M}$ on an assignment $a$ iff for some name $\beta$ in the language, $\phi[\beta/\alpha]$ (the result of substituting $\beta$ for every free $\alpha$ in $\phi$) is true in $\mathcal{M}$ on $a$.

$\ulcorner \Pi\alpha\phi \urcorner$ is true in $\mathcal{M}$ on an assignment $a$ iff for every name $\beta$ in the language, $\phi[\beta/\alpha]$ (the result of substituting $\beta$ for every free $\alpha$ in $\phi$) is true in $\mathcal{M}$ on $a$.

When it comes to proofs, the rules for substitutional quantifiers are exactly the same as those for objectual quantifiers. Even though the different kinds of quantifiers have different meanings, they are governed by the same rules of inference. (How can that be?)

What is substitutional quantification good for? Several interesting applications have been explored in the philosophical literature.

## 3 Nonexistence claims

As Linksy notes ("Two Concepts of Quantification," p. 227), statements of nonexistence have always been a bit tricky for standard theory of quantification. The following inference raises a puzzle, because its premise seems true, its conclusion false, and its form valid:

(1) Pegasus does not exist.

(2) $\exists x(x$ does not exist)

Suppose that the quantifier in (2) is understood in the standard way, as an objectual quantifier. Then the truth of (2) requires that there be an object in the domain that makes '$x$ does not exist' true when assigned to $x$. But this can only happen if the domain includes *nonexistent objects.* One way of resolving our difficulty, then, is to countenance nonexistent objects in our domain of quantification. (This approach is often associated with the philosopher Meinong. Quine raises some difficulties for it in his classic article "On What There Is.")

Another approach would be to deny that (1) is true. We might say: 'Pegasus' does not refer to anything, so we can't use it in a sentence to make a determinate claim that is capable of being true or false. (What, exactly, would we be talking about and calling nonexistent?) This is a hard line to take, since it certainly seems that we are saying something true when we use (1).

A third approach would be to deny that the argument has a valid form. The most obvious way to represent in in first-order logic is as follows (defining '$x$ exists' as '$\exists y(y = x)$'):

| | | |
|---|---|---|
| 1 | $\neg\exists y(y = p)$ | |
| 2 | $\exists x\neg\exists y(y = x)$ | ∃ Intro 1 p/x |

So construed, the argument has a valid form. But it might be argued that the form of (1) is really

$$\neg\exists x Px$$

for some *predicate P* (following Quine, we might use the verb "pegasizes"). On this approach, we say that (1) really means "nothing pegasizes," that is, there is nothing that has the properties traditionally associated with Pegasus (wings, horsiness, etc.). If that's right, then the argument is not valid, because it has the form:

| | | |
|---|---|---|
| 1 | $\neg\exists x Px$ | |
| 2 | $\exists x\neg\exists y(y = x)$ | ??? |

A fourth approach is to handle the problem with substitutional quantification. Surely, we might think, there is no assignment of values from the domain to $x$ that makes '$x$ does not exist' true. But it is much more plausible to think that there is a substitution instance of '$x$ does not exist' that is true. 'Pegasus does not exist' is a plausible example. This is enough to make '$\Sigma x(x$ does not exist)' true. So, if we understand the form of the argument as

| | | |
|---|---|---|
| 1 | $\neg\exists y(y = p)$ | |
| 2 | $\Sigma x\neg\exists y(y = x)$ | Σ Intro |

we can see it as valid without rejecting the premise (as the second approach does) or accepting that there are nonexistent objects in the domain (as the first approach does).

However, there's a catch. If we do this, we need to find a way of understanding the formula '$\neg\exists y(y = p)$' on which it is true. If '=' has its normal meaning, then '$(y = p)$' can only be evaluated for truth or falsity (relative to a model and assignment) if '$y$' and '$p$' denote something (relative to that model and assignment). But if '$p$' denotes anything in the domain, '$\neg\exists y(y = p)$' will be false. We might say that '$p$' denotes something *outside* of the domain—but now aren't we back to talking about nonexistent objects in our semantics? Alternatively, we might give truth conditions for '$y = p$' in some completely different way. But how? And how do we know when to use this other way and when to use the normal way?

*Exercise:*

3.1 Which of these four approaches do you think is best, and why? Write up your thoughts in a page or so.

## 4 Quantifying into attitude constructions

This is an issue we'll think more about later, so this is just a taste. Suppose we think that

(7) Caesar believed that Juno favored him.

It's natural to want to existentially generalize:

(8) There's a goddess who Caesar believed favored him.

But we might also want to say:

(9) There's no Juno. (Or: Juno does not exist.)

This is really tricky if we take the quantifiers involved to be regular, objectual quantifiers. Even trickier are cases like the following:

(10) Sarah thinks Eminem is clever.

(11) Sarah does not think Marshall Mathers is clever.

Now, we know that

(12) Eminem is Marshall Mathers.

So what should we say about this?

(13) There is someone who Sarah thinks is clever.

This sentence is true if there's an object in our domain that has the property of being thought clever by Sarah. And of course, there is such an object—Eminem. But wait: Eminem *is* Marshall Mathers, and it's not the case that Sarah thinks *he's* clever.

And what about this apparent consequence of (10) and (11)?

(14) There is someone, $x$, such that Sarah thinks $x$ is clever and Sarah does not think $x$ is clever.

How can any element of our domain have both the (contradictory) properties of being thought clever by Sarah and *not* being thought clever by Sarah?

Like the last puzzle, this one has inspired several very different kinds of solutions. As before, substitutional quantification looks like it might help. If we accept (12), then we can't regard the occurences of 'Eminem' and 'Marshall Mather' in (10) and (11) as regular referential uses of names, since then = Elim would be licensed and we could derive a contradiction. This rules out treating the quantifier in (13) as a regular objectual quantifier. But there is no obstacle to treating it as a substitutional quantifier. We can do that as long as something settles the truth values of its substitution instances (for example, 'Sarah thinks Ken is clever', 'Sarah thinks Chen is clever', 'Sarah thinks Sarah is clever', etc.).

## 5 Sentence quantifiers

So far, we have only considered quantifiers whose variables occur in places that could be occupied by names. We'll say that a variable is in *name position* when the result of substituting a name for (free occurrences of) the variable would be grammatical. It's natural to try to make sense of quantifiers that bind variables in other grammatical categories. For example, in

(15) $\exists p(p \supset \bot)$

the bound variable occupies *sentence position*; it occurs in a place where a sentence or formula could go. And in

(16) $\exists F(Fa \wedge Fb)$

the bound variable occupies *predicate position*; it occurs in a place where a one-place predicate could go. We can even imagine a quantifier that binds variables in *connective position*!

(17) $\exists * ((A * B) \equiv (B * A))$

It is easy crucial to keep in mind the distinction between *being in sentence position* and *ranging over sentences*. A quantifier that binds variables in name position can range over sentences. For example, if I say

(18) Every sentence I have written is convoluted,

I have quantified over sentences, but the quantifier I have used binds variables in name position:

(19) $\forall x((Sx \wedge Wix) \supset Cx)$

So what would quantifiers that bind variables in sentence position look like? It seems doubtful that we have any such quantifiers in natural languages, so we have to try to understand a partly symbolic example:

(20) For all $S$, if the senator says that $S$ then $\neg S$.

Here it's clear that $S$ is in sentence position. Only a sentence can grammatically fit after a $\neg$, or after 'says that'. Verify this by substituting some names in these places.

*Exercise:*

5.1 How would you express what is said by (20) in English?

## 6 Quantifying into quotes

Suppose you wanted to say,

(21) For some $x$, Bush said, '$x$ is the worst threat to our security' when $x$ was already dead.

The underlying thought seems intelligible. But there are some problems. One problem is that the second '$x$' is in quotes, and Bush rarely uses variable names in his speeches. We already know how to deal with *that* problem:

(22) For some (name) $x$, Bush said, ⌜$x$ is the worst threat to our security⌝ ....

But what about that last bit, 'when $x$ was already dead'? In (22), $x$ ranges over *names*—at any rate, it is an assignment of a *name*, a piece of language, to $x$ that will make

(23) Bush said, ⌜$x$ is the worst threat to our security⌝

true. But names can't be alive or dead. For the last part of the sentence, then, we need '$x$' to range over people, not names.

Is there any way to clean this up, or are we just hopelessly confused in thinking that there was an intelligible thought to express in (21)?

Well, if we construe the quantifier substitutionally, we can make good sense of our sentence:

(24) $\Sigma x$ (Bush said, '$x$ is the worst threat to our security' when $x$ was already dead)

This is true just in case there is a name in our language that, when put into the slots below, makes a true sentence:

(25) Bush said, '__ is the worst threat to our security' when __ was already dead.

The fact that one of these slots is inside quotes doesn't matter in the least to the substitutional quantifier.

## 7 Defining truth

Tarski noted that any good definition of truth (for sentences) ought to imply all instances of

(26) '$S$' is true iff $S$

where $S$ is a sentence. He then proceeded to show how to give such a definition, restricted to a particular language, using considerable ingenuity and some set theory. (His work on defining truth is the basis of the definitions of truth in a model we have been considering.)

But, one might wonder, isn't it just *trivial* to give a definition that meets Tarski's criteria?

(27) $\forall S(S \text{ is true} \equiv S)$

You can see the problem, I hope. If not, ask yourself: to what grammatical category does the bound variable $S$ belong? Is it in name position or sentence position? Ah yes—there's the rub. It needs to be in both, and that's impossible.

An obvious fix is to put the first occurrence of '$S$' inside quotes, since a sentence inside quotes is a *name* for a sentence, and thus the right kind of thing to go in front of 'is true':

(28) $\forall S(\text{'}S\text{' is true} \equiv S)$

But now we face a different problem. The quantifier can't penetrate inside the quotes, so we end up just talking about the letter '*S*'.

Could we fix this problem with corner quotes? Well, let's see:

(29) $\forall S(\ulcorner S \urcorner \text{ is true} \equiv S)$

The trouble is that '$\ulcorner S \urcorner$' is just the same as '*S*' by itself; both denote the expression denoted by '*S*'. So (29) is really no different from (27). The problem is the same: '$\ulcorner S \urcorner$' is in name position, so we still have a mismatch with the '*S*' at the end of the sentence, which is in sentence position.

What we need is clearly substitutional quantification:

(30) $\Pi S(\text{`}S\text{'} \text{ is true} \equiv S)$

However, as Linsky points out, this isn't general enough; it shows us how to deal with 'is true' when it applies to a quote name of a sentence, but not when it applies to some other term denoting a sentence, like 'the third sentence on this page'. The fix is easy enough, though:

(31) $\forall x(x \text{ is true} \equiv \Sigma p(x = \text{`}p\text{'} \wedge p))$

Now we're getting somewhere!

Alas, there are two rather serious problems for the project of defining truth this way.

### 7.1 Quantifying into quotes and paradox

The first is that substitutional quantification into quotes leads to paradox. The argument can be found in Tarski's "The Concept of Truth in Formalized Languages"[1] and in Linsky's article. Let me quote Tarski's words in full:

> Let the symbol '*c*' be a typographical abbreviation of the expression '*the sentence printed on this page, line 6 from the top*'. We consider the following statement:
>
> > *for all p, if c is identical with the sentence 'p', then not p* [Note: This is, in fact, the sentence printed on the sixth line from the top of the page.]
>
> (if we accept [31] as a definition of truth, then the above statement asserts that *c* is not a true sentence).
>
> We establish empirically:
>
> ($\alpha$) *the sentence 'for all p, if c is identical with the sentence 'p', then not p' is identical with c.*
>
> In addition we make only a single supplementary assumption which concerns the quotation-function and seems to raise no doubts:
>
> ($\beta$) *for all p and q, if the sentence 'p' is identical with the sentence 'q', then p if and only if q.*
>
> By means of elementary logical laws we can easily derive a contradiction from the premises ($\alpha$) and ($\beta$).

Let us try to reconstruct Tarski's reasoning in our notation. Let *c* denote the sentence

[1] In his *Logic, Semantics, Metamathematics*, 161–2.

($c$) $\Pi p(c = \text{‘}p\text{’} \supset \neg p)$.

Then, Tarski claims, the following sentence should be true, since it just says what $c$ is:

($\alpha$) $c = \text{‘}\Pi p(c = \text{‘}p\text{’} \supset \neg p)\text{’}$.

He then reasons as follows (I presume). Suppose $c$ is true. Then instantiating ‘$p$’ in $c$ with $c$ itself, we get

(1) $c = \text{‘}\Pi p(c = \text{‘}p\text{’} \supset \neg p)\text{’} \supset \neg\Pi p(c = \text{‘}p\text{’} \supset \neg p)$.

But the antecedent here is just ($\alpha$), so by modus ponens we get

(2) $\neg\Pi p(c = \text{‘}p\text{’} \supset \neg p)$.

But this is just the negation of $c$. So if $c$ is true then $c$ is false. Now suppose $c$ is false, that is,

(3) $\Sigma p(c = \text{‘}p\text{’} \wedge p)$.

Then there is some sentence $p$ such that

(4) $c = \text{‘}p\text{’} \wedge p$.

But then it follows that

(5) $\text{‘}p\text{’} = \text{‘}\Pi p(c = \text{‘}p\text{’} \supset \neg p)\text{’}$,

so that (by ($\beta$))

(6) $p \equiv \Pi p(c = \text{‘}p\text{’} \supset \neg p)$.

But from (4) we can get (by $\wedge$ Elim)

(7) $p$,

so we can conclude

(8) $\Pi p(c = \text{‘}p\text{’} \supset \neg p)$,

which is of course $c$. So, if $c$ is false, then it's true, and if it's true, then it's false. Antinomy.

I am fairly sure that this is what Tarski had in mind. Scott Soames points out[2] that Tarski assumes at step (1) that the sentence-variables can be instantiated with sentences that themselves contain sentence-variables (like $c$). If we understand the sentence-variables as belonging to the metalanguage and as ranging only over sentences of the object language, the derivation of the paradox does not go through. Soames concludes that Tarski's objection to quantifying into quotes is misplaced, and that a definition along the lines of (31) could satisfy all of Tarski's conditions on a successful definition of truth.[3] Of course, one might have other reasons for doubting that truth can be defined in terms of substitutional quantification into sentence position.

[2]In the Appendix to his chapter "Tarski's Definition of Truth" in *Understanding Truth*.

[3]Soames' approach derives from a well-known article of Saul Kripke, "Is There a Problem about Substitutional Quantification," in Gareth Evans and John McDowell, eds., *Truth and Meaning*, Oxford University Press, 1976; see esp. pages 366–8.

### 7.2 The circularity worry

Even if the definition of truth using substitutional quantifiers can be made coherent, one might worry that it is somehow circular. Here's how the objection might go:

> If we interpret the sentence position quantifier in 'for some $p$, $a$ = '$p$' and $p$' substitutionally, the sentence has the following sense: there is a sentence which, when it replaces the two occurrences of '$p$' in the embedded formula, produces a *true* sentence. But our grasp of this sense presupposes an understanding of truth, so we cannot use this kind of substitutional quantifier to *define* truth—at least not if we hope to define truth in nonsemantic terms. As Mark Platts puts it in *Ways of Meaning*: "The problem is clear: substitutional quantification is defined in terms of truth, and so cannot itself be used to define truth" (quoted in Soames, p. 91).

Soames argues that this objection "embodies a subtle but fundamental mistake about the nature of substitutional quantification." He argues that although 'for some $p$, $a$ = '$p$' and $p$' is *true* if and only if there is a sentence which, when it replaces the two occurrences of '$p$' in the embedded formula, produces a *true* sentence, it does not *mean* that there is a sentence which, when it replaces the two occurrences of '$p$' in the embedded formula, produces a *true* sentence. According to Soames, "the proposition that is expressed by the [substitutional] quantification is not a metalinguistic proposition about expressions at all. To suppose otherwise is to confuse substitutional quantification with objectual quantification over expressions" (p. 91). He gives two arguments.

1. The sentence '$\Sigma n(n$ is hot$)$' does not express a metalinguistic proposition about sentences. For one can believe that for some $n$, $n$ is hot without believing anything about sentences. Similarly, the proposition this sentence expresses would have been true even if the word 'hot' had meant something different, but the metalinguistic proposition might not have been true in that case. [Perhaps the idea is that the proposition expressed by '$\Sigma n(n$ is hot$)$' is the (possibly infinite) disjunction of the propositions expressed by each sentence ⌜$N$ is hot⌝, where $N$ is a closed term in English.]

2. We define '$\wedge$' by saying that ⌜$(A \wedge B)$⌝ is true if and only if $A$ is true and $B$ is true. (This is the truth-table definition.) But this does not mean that 'Joe swims $\wedge$ Mary bikes' expresses a metalinguistic proposition about the truth of the sentences 'Joe swims' and 'Mary bikes'. So the fact that we explain the substitutional quantifier in terms of truth does not make it any less appropriate for use in a definition of truth than conjunction is.

In his short article "Why I Don't Understand Substitutional Quantification,"[4] Peter van Inwagen complains:

> If I could understand the sentence
>
> $S$ $\quad \Sigma x(x$ is a dog$)$
>
> then I could understand substitutional quantification. But I cannot understand this sentence. I cannot understand it because I do not know what proposition it expresses.

Proponents of substitutional quantification say that $S$ is true just in case some sentence that is the result of substituting a term for '$x$' in '$x$ is a dog' is true. But they deny that $S$ *says that* some sentence that is the result... Van Inwagen's complaint is that they don't say what $S$ *does* say.

The issues here are interesting and subtle, and it is worth thinking hard about Platts's objection, Soames's response, and van Inwagen's complaint. (A possible paper topic?)

[4] *Philosophical Studies* 39 (1981), 281–5.