The Link between Misinterpretation, Intentionality, and Mental Agency in the Natural Language Interpretation of "Fake"

Janek Guerrini

Abstract. In formal semantics of natural language, an intersective interpretation works for many adjectives: x is a French lawyer iff x ∈ {x: x is French} ⋂ {x: x is a lawyer}. For those adjectives for which this does not work, like "excellent", we still have, at worst, a subsective modification ({x: x is an excellent violinist} ⊂ {x:x is a violinist}). Neither of these applies to "fake", whose formal interpretation is a traditional challenge. In this paper, I propose an analysis of the semantics of "fake" in which the speaker's attribution of intentionality (derived or original) to the object or person of which she predicates fakeness is central. In fact, the boundaries between the properties that 'fake' modifies and those it leaves unchanged are moved in function of this attribution of intentionality. In a famous 1994 paper, Dretske argues that for something to be specifically mental it does not merely need to exhibit original intentionality. It also has to be capable of misrepresentation, i.e. be a structure having a content independent of its causes. I argue that this intuition is implicitly contained in the natural language use of "fake".

Author. Janek Guerrini, janek.guerrini@ens.fr.

1   The Problem

Nouns and adjectives are commonly described in the semantic type-theoretical approach as functions of type <e,t> that take an element or an individual e, and return a truth-value t. For example, "lawyer" is a function that says "give me an individual, and I'll tell you whether or not it's a lawyer." These functions are the characteristic functions of sets, so that properties in the traditional sense can be seen interchangeably as functions and as sets of individuals. What happens when I combine two nouns, or two adjectives, or a noun and an adjective?

Consider sentence (1)

(1) x is a French lawyer

which says x ∈ {French}∩{lawyer}. In fact, "French" is an intersective adjective. in that (A) holds for any N.

(A) ||French N|| = ||French|| ∩ ||N||

This approach seems a very elegant and obvious account of the compositionality of nouns and adjectives. Unfortunately, it works only for a portion of adjectives. Some cases for which it does not work are adjectives like "fake" or nouns like "toy". A fake violinist is not someone who is both fake *and* a violinist.

The problem can be stated as follows. A formal account of "fake" needs to be able to tell us why and how a "fake gun" is neither a gun nor merely a non-gun. Notice, in fact, that (2) is well-formed.

(2) That gun is a fake gun.

2   A first sketch of the account

We introduce an interpretation of nouns and adjectives as structured sets of properties.

||G|| = ⟨R,P⟩

Where:

- G is a noun or adjective.

- R is the set of all *relevant* properties of G that all instances of G must have.

- P is the set of all *prototypical* properties of G that all prototypical instances of G have. We assume R⊆P.

It is central to point out that R and P are not structured sets of properties themselves, but plain sets of properties. In this approach, properties are still the standard mathematical object: sets of individuals that have a given property.

Only the meanings of natural language words like adjectives and nouns are not merely properties, but structured sets of properties.

A basic example for the noun "gun":

1. Gun

    R= {shoots, kills}

    P= {has the physical form, has a barrel, shoots, kills}

Now we can give an analysis of 'fake':

f($\langle$R,P$\rangle$) := $\langle\{$ seems to r but cannot r:r$\in$R$\},\{P-R\}\rangle$

That is, f is a function that takes as an input a *structured set of properties*, both the relevant and the prototypical properties, and returns another structured set of properties, where

- every prototypical property is left unchanged

- every relevant property r is modified from "r" to "seems to r and cannot r".

A basic example:

1. fake gun = f(gun)

    R = seems to {shoot, kill}and cannot {shoot, kill}

    P = { has the physical form, has a barrel}

3   Discussion

3.1   3.1 Instability across contexts

I am not arguing that anytime we pronounce "fake gun" we intend the exact interpretation given above. Rather, if the context makes shooting and killing relevant to being a gun, 'fake' composes by taking only those properties that are relevant as input. But these R-properties change over different contexts, as showed by the example below.

Imagine a world in which guns, instead of keys, are used to open doors. Every gun has unique bullets, you shoot on your door, the door reads the bullet, and if the door recognizes it as its specific bullet it opens. These guns can still shoot and kill. Now I tell you: that's a fake gun.

All I am saying with this utterance is that it opens no door, and not that it is not able to kill. I am using a different R set than I would be using to say, in our actual world, something like "The gun he used to rob the bank was fake". In the gun-key world, the property of being able to kill might just be a prototypical property and not a relevant one. *How* the properties which are relevant to

a predication change over different contexts is not the subject of this paper. By contrast, I have shown that those properties change, and this is sufficient to propose an account of how 'fake' composes with heads by assuming it only applies to those relevant properties.

The fact not all adjectives have an intersective or subsective meaning, like "French" in a phrase such a French lawyer, was observed by Reichenbach as early as the '40s. In the sentence «John is a slow rider», "what is said is not that John is slow in general but only that John is slow in his driving: thus the word 'slow' […] operates as a modifier of 'drive'".

Attempts to describe fake not only as a non-privative predication, but also as an operation on some internal structure of NPs started as soon as the '80s. Lakoff and Johnson's account of privative NPs is based on the idea that privative Adjs operate ranging over the complex internal semantic structure of terms, especially in the case of artifacts (Lakoff and Johnson 1980). In 'fake gun', for instance, they put forward an analysis which includes three representational dimensions: perceptual, functional, and genealogical.

Franks' account shows that despite their apparent vagueness in classification ("what is a gun?"), NPs work in a quasi-classical way: their internal attribute-value structure determines uncertain predications, but the single attributes ('features') work in a perfectly binary fashion (Franks 1995). Franks' account is the first that posits a complex internal structure that distinguishes between central and diagnostic features. 'Fake' negates only the central features, keeping the diagnostic ones. This makes him predict that the central features of a gun do not obtain of a gun, but the gun does not seem to possess these, making that account of privatives significantly different from this account of 'fake'.

Partee (1987), by contrast, did not posit any rich internal structure, providing a non-intersective extensional account in which adjectives like 'fake' coerce their argument to a broadened extension. Thus in (2) «the first occurrence of gun, modified by fake, is coerced, whereas the second, unmodified, occurrence is not. Normally, in the absence of a modifier like fake or real, all guns are understood to be real guns, as is evident when one asks how many guns the law permits each person to own, for instance. Without the coerced expansion of the denotation of the noun, not only would fake be privative, but the adjective real would always be redundant. »

More recently, Del Pinal (2015) proposed another kind of internal complex structure on which 'fake' acts. NPs are constructed as having one principal extension-determining, competence-linked sub-structure and another, sub-structure containing all the core facts about that noun, secondary and normally not involved in compositionality. 'Fake' works differently than other modifiers because it takes as an input the Cstructure. Hence a fake gun appears to fulfill all the core-facts we know about a gun. Such an extensional semantics can be

employed in a HeimKratzer-like, traditional formal semantics.

This need of positing a rich structure, but still directly extension-linked (without intensional mediation), comes as a result of considering Putnam's influential argument against definitional theories (1970). The argument goes as follows: for any property that supposedly defines an artifact, we can always find a counterfactual situation in which an object falls under the extension linked to that artifact in spite of not having that property. Therefore, words are linked directly to the extension they denote, and it is only through word-use that we understand exactly what individuals fall under an extension.

This helps us address the issue of intensionality. Clearly this is not the topic of the paper: the proposed account presupposes an intensional structure, but it also works by partitioning an extensional space, hence embracing a view sympathetic with Putnam (and seeing the properties we posited above as sets of individuals). However, I want to state my sympathy for intensional internal structures. I want to limit myself to observe that if reference and compositionality occur on a part of the features of an internal structure (as is the case in R vs P), and if the bundle of relevant features is not stable across contexts, then Putnam's argument is less effective. Of course I can think of a counterfactual situation in which an individual not having the property of 'being a hunting big cat with a mane" falls under the extension of lions. But this is because in that situation the properties relevant to this categorization are different.

3.2   Improving the analysis

This is why I did not call R.properties essential. An object's essential property, if it exists at all, is possible world-invariant and would have to hold, in virtue of this, across conversational contexts. For instance, if being a mammal is an essential property of humans, then it is necessarily an essential property of humans. Fake objects, by contrast, enjoy bigger flexibility in that what the gun is made for changes depending on the possible worlds that are being referred to. It looks like, at least in the case of tools, relevant properties amount to the purpose the object was made for. And there are good reasons to make an even stronger claim, namely that some intentionality in the meaning of fake or in the object to which we apply the function f plays a big role in the interpretation of the whole. The following thought experiment should clarify this point:

> Bob and Carl have built a super-powerful telescope: not only can it show you the furthest galaxies, but you can also look through the objects you are pointing it to. By playing around with the telescope and looking at an ice-cold, lifeless and extremely distant galaxy, Bob happens to point at an atom-conglomerate which randomly appears exactly like a gun. Bob calls Carl and they observe it together. How-

ever, by looking inside the gun with the distance-rays-function of their powerful telescope, Bob and Carl, who are also weapon-freaks and know a lot about how weapons work, notice that, given how the internal mechanism works, the gun could never shoot and kill. Everything else resembles exactly a gun.

In this situation, it wouldn't be appropriate to say: "look, a fake gun!", for the only way in which this can become felicitous is by cooking up a context that posits a builder and a purpose assigned to the gun.

If this is the case, then "fake" implies at least some kind of derived intentionality. By intentionality here I intend the standard Brentano definition (Brentano 1874). Brentano referred to intentionality as the power of minds to be about, to represent, or to stand for, things, properties, and states of affairs. In philosophy of mind debates this is nowadays referred to as "original" or "primitive" intentionality. "Derived" intentionality is borrowed from original intentionality: a gun borrows its intentionality from its conceiver. A gun is about shooting and killing, a hammer is about beating nails.

One could with good reason contest that it is perfectly fine and reasonable to talk about a fake lawyer and that this lawyer was not created with the purpose of seeming a lawyer. But this does not change the fact that when stating a sentence of the form

(3) x is a fake G

you are implying G has some kind of intentionality. This intentionality may be derived, as for the gun, or original, as for the lawyer.

But the example with Bob and Carl leaves us with a question about the account of "fake" that we put forward above. The conglomerate of atoms seems to shoot, cannot shoot and has all of the prototypical properties of a gun, including the physical form , and nevertheless cannot be said to be a fake gun.. Yet the account I proposed above predicts it to be a fake gun. What is missing?

There are two ways to restrict the account as a means of making the right predictions:

I. Not only does "fake" change relevant properties to "seems to R and cannot R", but it crucially **adds** that it was built by its builder with the purpose of seeming R.

II. "fake" as we defined it is fine. We just have to specify that it applies only to things that already have a purpose, a derived intentionality. That is why it is semantically awkward to say something like "look, a fake atom!", modulo the possibility of pragmatic adjustments that posit a builder of that atom. It is really not clear what this should mean and it does not at all look like something a competent speaker

would say. The sentence above is therefore unclear unless we try to think out a story in whichthat atom was built by someone who deliberately put into it the purpose of resembling through and through an atom. In this case we would no longer be talking about semantics as it would be a pragmatic adjustment.

Basically, our decision must be whether to add some operations to what the function "fake" already does (I.) or to restrict its domain (II.).

Before making the decision, it may come in handy to unpack what exactly "seem" means when we are saying that a fake G seems to R and cannot R. A reasonable account for sentences like

(4) j seems to be playing

might be that when you utter such a statement you are saying that you think that an observer will think that j is playing. Now we can unfold a more complete interpretation of "fake":

$f(\langle R,P \rangle) := \langle \{$ is capable of fooling j into thinking that r and cannot r:r$\in$R$\}, \{P - R\} \rangle$

Where j is a particular observer, a 'judge' or, in its absence, an average observer. This intuition is drawn from the Moltmann 2010 account of taste predicates. Consider the two following examples:

(a) Ron's eyes are shaped in a particular manner, which cannot see the difference between a gun and a drill. Dan usually threatens him with a drill. When he tells the story to his brother, he says "And then I was pointing my fake gun towards him, and…"

(b) "Monopoly bills are fake".

In case (a), j is Ron. In case (b), j is clearly the average observer modulo the assumption of a lack of knowledge about the nature of monopoly.

To translate this into an example, postulating that in a given context, i.e. in a given possible-worlds set

3 lawyer

   R= {has a law degree, is member of the bar}

   P= {wears a robe, has good rhetoric}

4 fake lawyer=f(lawyer)

   R= is capable of fooling j into thinking that{has a law degree, is member of the bar} and it is not the case that {has a law degree, is member of the bar}

   P= {wears a robe, has good rhetoric}

The judge parameter is, alongside with the intensional structure, what differentiates this analysis from Del Pinal's, which states that a fake gun was made to have the perceptual features that make it look like a gun, but doesn't make any predictions regarding to whom the fake gun should look like a gun.

We still have to decide whether "fake" not only modifies but also adds some operation to the function it applies to (I.), or it has just a domain restricted to what the speaker attributes some derived or original intentionality (II.). It is very unclear which of the two options delivers a better natural language meaning. The different predictions made by these two hypotheses concern whether the mental-agent-requiring content is at-issue or presuppositional. At-issue content is plainly relevant to truth conditions. So if the at-issue content of a sentence is false, then its negation is true. Presuppositional content, by contrast, works differently. Take for instance the famous Russellian example "The king of France is bald". Because the presupposition of the sentence is not fulfilled, it is neither true nor false, and the same hold for its negation. A characteristic of presuppositions is that, unlike at-issue content, they project out of embedded clauses. Take following sentences:

(i) If that thing over there is a fake atom, I'll be really surprised.

(ii) That thing over there was built by someone, that's for sure, but if it's a fake atom I'll be really surprised.

Under hypothesis I sentence (i) presupposes that "that thing over there was built by someone," while (ii) triggers no presuppositional content. In fact, the content that is supposedly presupposed in (i) is asserted in (ii), and thereby made at-issue. Under hypothesis II neither sentence presupposes anything. Another example that might help clarify is following:

Under hypothesis I sentence (i) presupposes that "that thing over there was built by someone," while (ii) triggers no presuppositional content. In fact, the content that is supposedly presupposed in (i) is asserted in (ii), and thereby made at-issue. Under hypothesis II neither sentence presupposes anything. Another example that might help clarify is following:

A: Oh my gosh, there's a fake atom over there!

B: What you just said is false, since no one built that thing.

B: What you just said is false, since no one built that thing.

Again, if B' seems more of a plausible reaction than B, then it makes a (all but uncontroversial) case for II. So all really depends on which intuitions we have: is there (more of) a presupposition failure in (i) there than in (ii)? Is B' more plausible than B? An informal survey among my colleagues found contrastive intuitions. Experimental work might be needed on this, as judgements are very

subtle here. Del Pinal, on the other hand, seems to be more empathetic with I, as he makes the way in which a thing comes into being at-issue. If we were to make the same choice, we would have following meaning assigned to 'fake':

f(⟨R,P⟩) := ⟨{ was built with /has the precise purpose of fooling j into thinking that r and cannot r:r∈R},{$P - R$}⟩

3.3   Intentionality and misrepresentation

In his 1994 paper, Dretske argues that intentionality per se is not mental. For take the example of a compass:

(a) the compass indicates the North-pole

(b) the North-pole is coextensional to the habitat of polar bears

The conclusion of

the compass indicates the habitat of polar-bears

is not justified. We have generated an intensional and referentially opaque context. This is enough in order for something to have original intentionality. Dretske defines intentionality in the following terms:

if ascribing a property to x generates an intensional context, then x exhibits original intentionality.

We already defined derivative intentionality as pertinent to those objects (like a gun) whose representing such-and-such can be explained in terms of the intentionality of something else. Then how is a compass different from a gun?

The intentionality of the device is not, like the intentionality of words and maps, borrowed or derived from the intentionality (purposes, attitudes, knowledge) of its users. The power of this instrument to indicate north to or for us may depend on our taking it to be a reliable indicator (and, thus, on what we believe or know about it). but its being a reliable indicator does not itself depend on us.

As the compass de facto exhibits original intentionality, original intentionality cannot be the distinguishing mark of thought. So mere intentionality involved in the semantics of fake wouldn't be enough to prove that we represent any mental agent behind 'fake'. But what artefacts, however sophisticated, cannot do, argues Dretske, is misrepresenting something without our help:

Although clocks, compasses, thermometers, and fire alarms-all readily available at the comer hardware store-can misrepresent the conditions they are designed to deliver information about, they need our

help to do it. Their representational successes and failures are underwritten by and, therefore, depend on our purposes and attitudes, the purposes and attitudes of their designers and users. As representational devices, as devices exhibiting a causally detached meaning, such instruments are not therefore eligible ingredients in a recipe for making thought. (Dretske 1994)

Now: "fake" always hides mental agents with misrepresentational states behind its meaning, for note that:

- we showed that we assume of a hypothetic average observer that he would be fooled by the object / person. To be fooled, the observer must form himself a misrepresentation. In other words, there must be some semantic content independent of its causes.

- we attribute to the builder from which the object borrows intentionality or to the person of which we predicate fakeness the active misrepresentation of the relevant properties of the thing of which they are a fake version.

Thus

(a) Misrepresentation needs a specifically mental

(b) The adjective "fake" needs misrepresentation

(c) we always need a mental agent or an intentionality derived from a specifically mental agent when we use "fake".

Note indeed that mere derived intentionality is not enough for fake objects. For instance, if I want to build a paperweight but by pure accident end up making something that resembles a gun, is that a "fake gun"? No, and yet it definitely has derived intentionality. I built it with a purpose (being a paperweight) that it can accomplish. What is needed behind the fake object is an *active misrepresenter who has the goal of that exact misrepresentation.*

In conclusion, I turn to explaining why (2) is a perfectly fine sentence.

(2) That gun is fake.

That object can be described as a "gun" only from the point of view of the person who is fooled, and as "fake" only from the point of view of a fully informed person. But we know that there are observers that represent that thing as a gun. It seems that there is a perspective shift between "gun" and "fake" taking place in (2), from the fooled person to the speaker (or from the average person to the well-informed speaker in case of a the standard filling the void argument). The speaker says "gun" by putting himself in the average observer's frame, and then switches to his better-informed frame in order to predicate "fakeness" of the object.

By shifting frame of reference, we are shifting between different boundaries between R and P . The domain of conversation modifies these boundaries only insofar as it modifies what type of intentionality is behind of the objects of the conversation. We delimit R and P in function of the mental agents we have to posit in order to be able to predicate 'fakeness': the average observer and the lawyer or the builder of the gun. We understand each other and agree in the vast majority of occurrences in what respects an object is "fake". This shows how cognitively convenient it is for us to posit mental agents. How good we, as humans, are at understanding others' representational and misrepresentational states.

For this reason, accepting the analysis proposed in this paper and the linked view does not necessarily mean endorsing an intentional view of the mind rather than a computational one. Dennett himself Dennett 1987 concedes that intentionality is for humans a convenient way to think about others' mental states. Ascribing to our chess opponent psychological states with intentionality makes it easier for us to imagine what she is thinking, planning, avoiding, fearing etc. All one is committed to when accepting this analysis is that we do indeed ascribe a mental intentionality mainly characterized by capability of misrepresentation to others, not that the intrinsic nature of the mind itself is intentional.

Acknowledgements

References

Brentano, Franz (1874). *Psychology From an Empirical Standpoint*. Routledge 1995.

Coquand, Thierry (2015). *Type Theory*. Ed. by Edward N. Zalta. The Stanford Encyclopedia of Philosophy. URL: https://plato.stanford.edu/archives/sum2015/entries/type-theory/.

Del Pinal, Guillermo (2015). "Dual Content Semantics, privative adjectives, and dynamic compositionality". In: *Semantics and Pragmatics* 8.7, pp. 1–53.

Dennett, D. and J. Haugeland (1987). "Intentionality". In: *The Oxford Companion to the Mind*. Ed. by R.L. Gregory. Oxford: Oxford University Press, pp. 139–143.

Dennett, Daniel (1987). *The Intentional Stance*. Cambridge, Mass.: MIT Press.

Dretske, Fred (1994). "If You Can't Make One, You Don't Know How It Works". In: *Midwest Studies in Philosophy* 19.1, pp. 468–482.

— (1995). *Naturalizing the Mind*. Cambridge, Mass.: MIT Press.

Fodor, Jerry (1975). *The Language of Thought*. New York: Crowell.

Franks, Bradley (1995). "Sense generation: A quasi-classical approach to concepts and concept combination". In: *Cognitive Science* 19.4, pp. 441–505. URL: http://dx.doi.org/10.1207/s15516709cog1904_2.

Jacob, Pierre (2014). *Intentionality*. Ed. by Edward N. Zalta. The Stanford Encyclopedia of Philosophy. URL: https://plato.stanford.edu/archives/win2014/entries/intentionality/.

Klein, Ewan (1980). "A semantics for positive and comparative adjectives". In: *Linguistics and Philosophy* 4.1, pp. 1–45.

Lakoff, George and Mark Johnson (1980). *Metaphors we live by*. Chicago, IL: The University of Chicago Press.

Parsons, Terence (1970). "Some problems concerning the logic of grammatical modifiers". In: *Synthese* 21.1, pp. 320–334.

Partee, Barbara (1987). "Lexical Semantics and Compositionality". In: *An Invitation to Cognitive Science (Second Edition). Volume 1: Language*. Ed. by Lila Gleitman. Cambridge, Mass.: MIT Press.

— (2010). "Privative: Subsectives plus coercion?" In: *Presupposition and discourse: Essays offered to Hans Kamp*. Ed. by R. Baule, U. Reyle, and T.E. Zimmerman. Amsterdam: Elsevier.

Putnam, Hillary (1970). "Is semantics posible?" In: *Metaphilosophy* 1.3, pp. 187–201. URL: http://dx.doi.org/10.1111/j.1467-9973.1970.tb00602.x.