Pretense for the Complete Idiom[1]

Andy Egan

University of Michigan/Australian National University

Introduction

Idioms – expressions like *kick the bucket* and *let the cat out of the bag* – are strange. They behave in ways that ordinary multi-word expressions do not. One distinctive and troublesome feature of idioms is their *unpredictability*: The meanings of sentences in which idiomatic phrases occur are not the ones that we would get by applying the usual compositional rules to the usual meanings of their (apparent) constituents. This sort of behavior requires an explanation. I will argue that the right explanation is that the sentences are being interpreted through a pretense. (What this means, exactly, will be explained in what follows.) This is a surprising claim, for two reasons. On the one hand, it seems that adopting a pretense account is *overkill*—it's a far more radical move than is required to account for the phenomena. On the other hand, it seems that a pretense account is *hopeless*—that there is a fatal overgeneration problem for pretense accounts of idiom that causes them to fail "as badly as it is possible for an account of idiom to fail."[2] So I will have some work to do.

Before we get started, a word about our subject matter. I'm concerned with how we produce and understand sentences containing *idioms*, where the class of idioms is picked out by pointing at examples. So I don't want to define the class of idioms as all expressions meeting some condition C, but instead as all expressions of the same kind as *kick the bucket, let the cat out of the bag, take the bull by the horns, break the ice,* and so on. These

expressions form an interesting and oddly behaved class, but I don't want to prejudge anything by stipulating the features that an expression has to have in order to be a member of the class. (Note that the class of expressions that I'm concerned with here almost certainly will not include every expression that might, by some reasonable criterion or other, be classified as an idiom. Our everyday, pretheoretical use of "idiom" probably doesn't pick out a homogeneous natural kind. What I'm interested in doing here is providing an account of a natural kind to which some large-ish number of the expressions that fall under the everyday term belong.)

I should also emphasize that this is a paper primarily about *idioms*, which happens to use a pretense theory as a tool for explaining some of the peculiar behavior of these particular bits of natural language, and not a paper about *pretense in general* that happens to use idioms as an example. The questions that motivate the present paper are questions about how to account for our ability to understand and manipulate idioms in the ways that we do. I take a pretense theory to be helpful in providing such an account. I'm not (at least in this paper) concerned with any of the interesting and difficult general questions about the viability of pretense theories of any bit of language whatsoever, except insofar as these questions apply directly to the case of idioms. In particular, I will not be concerned with questions about the applications of pretense theories of various bits of language in the service of distinctively metaphysical goals, as means for avoiding unwanted ontological commitments.

In addition, though I'll be drawing parallels with metaphor, this won't be a paper about theories of metaphor. I *will* make some contentious assumptions about how metaphors work,

which I will not defend. In particular, I'm going to assume that a pretense theory of metaphor (of the sort advocated by, for example, Kendall Walton (1990, 1993, 2000)) is correct. Pretense theories of metaphor have been defended elsewhere (notably in Walton's work cited above), and any discussion I could add here, without making this paper intolerably long, would be unlikely to add anything helpful to that debate. (Though I do take myself to be, at a couple of points, providing *some* support for pretense accounts of metaphor, since many of the features of idioms that I will use to motivate a pretense account are also present in metaphors.)

With that in mind, let's have a look at the peculiarities of idioms that make them deserving of some sort of special treatment.

1. Why Idioms Are Weird

Idioms have two peculiar features. Here is the first:

> UNPREDICTABILITY: The meaning of a sentence in which an idiom occurs is
> different from the meaning you'd get by applying the usual compositional rules to
> the usual semantic values of its (apparent) constituents.

For example, you don't get the right meaning for "Livia let the cat out of the bag" by composing the usual meanings of *Livia*, *let*, *the*, *cat*, *out*, *of*, and *bag* in the usual way. That will get us a proposition that's true just in case Livia released some salient cat from some

salient sack, not one that's true just in case she revealed some salient secret, which is the one

that we want.

Another unusual feature of idioms is:

INFLEXIBILITY: Idioms are *frozen* in ways that other expressions are not.

Apparently innocent changes in wording or structure often make the idiomatic

reading of a sentence in which an idiom occurs unavailable, or at least strained.

Here are some examples that demonstrate the inflexibility of idiomatic constructions:

1. (passive)
   a. Tony blew off steam.
   b. *Steam was blown off by Tony.

2. (anaphora)
   a. Tony shot the breeze with Junior, and Paulie shot the breeze with Silvio.
   b. *Tony shot the breeze with Junior, and Paulie shot it with Silvio.

3. (substitution of synonyms)
   a. Richie kicked the bucket.
   b. *Richie kicked the pail.

It's worth saying a bit about the respect in which the starred sentences are deviant or

defective. It's not that the sentences are ungrammatical, or semantically uninterpretable, or

even that the idiomatic readings of the sentences are nonsensical or completely unavailable.

Rather, the problem is that there's a certain awkwardness to the idiomatic use. In addition, in cases where the literal reading isn't crazy or nonsensical, there'll probably be pressure to interpret the starred sentences literally rather than idiomatically – the audience might not smoothly and automatically latch on to the idiomatic meaning.

So idioms have two unusual features, which we'll need to do some theorizing in order to explain. In particular, given UNPREDICTABILITY, we'll have to do some theorizing in order to explain how we're able to use and understand idioms at all.

2. Three Theories of Idioms

UNPREDICTABILITY tells us that sentences containing idioms don't get their meanings in the usual, compositional way. The task, then, is to figure out how these sentences *do* get their meanings.

Very plausibly, what we should say is that *sentences* containing idioms get nonstandard interpretations because the *idiomatic phrases* (like *kick the bucket, pull strings* and *let the cat out of the bag*) get nonstandard interpretations. So "Tony pulled strings to keep Chris out of prison" gets its nonstandard interpretation because "pulled strings" gets a nonstandard interpretation.

The most straightforward version of this sort of view is the *lexical item view*. According to this view, idiomatic phrases are semantically unstructured. The apparent constituents of idiomatic phrases (like *kick* in *kick the bucket*, or *cat* in *let the cat out of the bag*)

aren't genuine constituents—they don't make any independent contribution to the meaning of the whole phrase. Idiomatic phrases are just extra lexical items, with their own semantic values given by the lexicon, rather than being inherited in the standard way from the semantic values of their apparent constituents. The caricature slogan is: *kicked the bucket* is just a funny way of pronouncing *died*. (That's a caricature because the view isn't that *kick the bucket* and *die* are the same lexical item, with different spellings and pronunciations, but that they're two different lexical items that make the same contribution to the truth conditions of sentences in which they occur.)

So the first candidate view of idioms is:

LI: Idiomatic phrases are semantically unstructured lexical items.

This has been a popular view. Some examples of those who have endorsed it are Jerrold Katz (1973), Noam Chomsky (1980), Martin Davies (1983), and Richard Moran (1997).

There is another way to hold a view according to which sentences containing idioms get nonstandard interpretations in virtue of the idiomatic phrases getting nonstandard interpretations. According to this view, it's the *parts* of an idiomatic phrase, not the phrase as a whole, that gets a nonstandard semantic value. These parts are then composed according to the usual compositional rules.

So in "Livia let the cat out of the bag", *the cat* and *let ___ out of the bag* get interpreted as (roughly) *the information* and *revealed*, respectively. Applying the usual compositional rules then gets us the right truth conditions for the sentence.

The second candidate view, then, is:

CHUNKS: Parts of idiomatic phrases have nonstandard semantic values when they occur within the idiom, which then get composed in the usual way.

This is the sort of theory proposed in Nunberg, Sag, and Wasow (1994).

The final option that I'll consider here is a pretense account. This sort of theory is different from both LI and CHUNKS in that it doesn't assign any nonstandard semantic values to any of the parts of idiomatic sentences. Instead, idiomatic phrases and their parts all have their usual semantic values, which all get composed in the usual way, but then something *else* happens that gives the sentence as a whole its nonstandard truth conditions. The *something else* is that the sentence is interpreted through a pretense. (It shouldn't be terribly clear exactly what this means yet—I'll spell out the pretense theory more carefully in the next section).

So the final candidate view that we'll consider here is:

PRETENSE: The parts of sentences containing idioms all retain their usual semantic values, and are composed in the usual way, but the sentence is assigned nonstandard truth-conditions by processing its literal content through a pretense.

As mentioned above, the notion of 'processing through a pretense' needs more explanation before we can really evaluate this sort of account. The task of the next section will be to spell out what PRETENSE amounts to in more detail.

3. Explaining PRETENSE[3]

We often engage in pretenses of one sort of another. There are, for example, make-believe games like mud pies and cops and robbers, and works of fiction like the Phillip Marlowe stories. Within these pretenses, there are certain things that are to be pretended (that are true according to the relevant fiction), and other things that are not.

When we go into the front yard and pretend that the cars on the street are buffaloes, not everything that's true in our newly instituted buffalo game (true according to the relevant fiction—henceforth *fictional*) needs to be explicitly pretended. Once the basic framework of the buffalo game is in place, the fact that I ran out into traffic (not merely the guiding cars-are-buffaloes fiction) makes it fictional that I've nearly been trampled by a buffalo, and the fact that the car that almost got me swerved and honked makes it fictional that the buffalo spooked and bellowed. Once we're pretending that clouds have faces, *the shape of the cloud* makes it

fictional that the cloud on the horizon is scowling. And it's (in part) the actual facts about American geography, not just what's written down in the Raymond Chandler stories, that makes it fictional that Phillip Marlowe lived 900 miles from Albuquerque.

Real-world facts have this capacity to give rise to truths in the fiction because pretenses are defined by a number of *principles of generation* that determine what's fictional. For example, the central principle of generation for the buffalo game is, *wherever there's a car, pretend that there's a buffalo*. Some likely further principles are, *if a car honks its horn, pretend that the buffalo bellowed*, and (for bad children) *if a BB hits a car, pretend that the corresponding buffalo was shot*. So we don't need to build it in to the buffalo-game fiction that there are seven buffalo in the street, and one of them is bellowing after being shot. We just introduce the principles of generation, and let the cars and the kid with the BB gun do the rest.

What's fictional, then, depends on two things: What the world is actually like, and which principles of generation are in effect. It's fictional that there's a stampede because (a) there are a lot of cars going by, and (b) the *where there's a car, pretend there's a buffalo* principle of generation is in effect. If fewer cars were going by, there would only be a fictional trickle of buffalo. And if we were pretending that cars were ostriches, there would be a fictional stampede of ostriches rather than buffalo.

So the fiction-independent facts about real-world objects can give rise to various fictional facts, as in the examples above. You can tell me things about what's true in the Marlowe fiction or the buffalo game by telling me things about 1930s California or passing

cars.

The important thing to notice for our purposes is that this is *reversible*. You can also tell me things about our actual California and real-world traffic patterns by telling me things about Phillip Marlowe and herds of buffalo. If you say, "Marlowe lived 350 miles from San Francisco", you tell me something about the *actual* distance between San Francisco and Los Angeles. When you say that the buffalo are just moseying by today, you tell me something about how fast the traffic is moving. So it's possible to use a fiction in order to make claims about the non-fictional world, because what's *fictional* depends in part on what's *actually true*.

When we say things about the real world by saying things about (or within) a fiction, what's said about the world depends on (a) what's said about (or within) the fiction, and (b) which principles of generation are in effect.[4] The same utterance, interpreted through different pretenses, might convey very different information about the real world. This is because very different real-world facts might deliver the same fictional truths, depending on which pretense we're engaging in (that is, depending on which principles of generation are in effect). It could be fictional that there's a stampede either because it's rush hour, and we're pretending that *cars* are buffalo, or because the elementary school just let out for the afternoon, and we're pretending that *fourth-graders* are buffalo. So depending on which pretense we're engaging in, when I say, "it's a stampede!" I could be telling you either that there are a lot of cars on the way or that there are a lot of fourth-graders on the way. The real-world significance of what's

true according to the fiction, as well as the fictional significance of what's actually true, depends on which principles of generation are in effect.

With this apparatus in place, I can give a better explanation of how a pretense account of idiom is supposed to work. Remember the central claim of the pretense account:

PRETENSE: The parts of sentences containing idioms all retain their usual semantic values, and are composed in the usual way, but the sentence is assigned nonstandard truth-conditions by processing its literal content through a pretense.

According to PRETENSE, each idiom has an associated pretense, and interpreting an idiom is a two-step process. First, we get the literal content of the sentence via the usual compositional process. This tells us what to pretend. Then we use the principles of generation to figure out what would have to *actually* be the case in order for the principles of generation to make the literal content fictional. This gives us the idiomatic truth conditions for the sentence.

For example, on such a theory, the principle of generation that pretty much defines the pretense associated with *kick the bucket* will be, *if somebody dies, pretend that there's some salient bucket that they kicked*. So when we go to interpret "Richie kicked the bucket", we get the literal content in the usual way, and then process that through the pretense. That is, we figure out what would have to *actually* be the case in order for the literal content to be *fictional*. In this case, given the principle of generation, what it takes for it to be *fictional* that Richie

kicked the bucket is for Richie *actually* to have died. So we get the right truth conditions: "Richie kicked the bucket", interpreted through the pretense, is true iff Richie died.

A number of people—most notably Kendall Walton (1993)—have advocated this, or something very like it, as an account of metaphor.[5] I think that they're right—that something like PRETENSE really is the right account of metaphor. In fact, I think that idioms and metaphors are interpreted in very much the same way, and that the distinction between them is not terribly sharp. I'll say a little bit more about this in the final section, but for the majority of the paper, I will studiously avert my eyes from metaphor in order to focus on the phenomena that motivate a pretense theory of *idioms*. Since many of the same phenomena are present in metaphor, there will be some parallel arguments that one could make in order to motivate a pretense theory of metaphor, but as this could very easily take us too far afield, I will not make them here.[6]

Now that we have PRETENSE more clearly in view, there are two fairly obvious complaints to make. The first is that this is overkill. It's using way too much fancy apparatus to do a fairly simple job. Why postulate this extra layer of processing, when we can explain the phenomena just by postulating a straightforward kind of ambiguity (either of idiomatic phrases or their components)? At this point in the discussion, this complaint is fair enough. If the only sentences we had to account for were like "Richie kicked the bucket", this really would be a far more elaborate story than was warranted by the phenomena. But in the next few

sections we'll see some phenomena that are harder to deal with, and which I take to justify the complicated machinery.

The second complaint is that this sort of account seems to be very badly suited to account for INFLEXIBILITY. If the truth conditions of idiomatic sentences are determined by (a) their literal content, and (b) the principles of generation that are in effect, then it's surprising that "Tony blew off steam" naturally receives an idiomatic reading while "steam was blown off by Tony" does not. After all, they have (plausibly) the same literal content, and the same principles of generation ought to be in effect, so they ought, if a pretense account is right, both to receive equally happy idiomatic interpretations, which deliver the same idiomatic truth conditions. But they don't. Section 7 will be devoted to solving this problem.

But let's shelve this complaint about PRETENSE for now, and look at some phenomena that make trouble for LI and CHUNKS. These problems will also provide our response to the overkill objection—the reason why PRETENSE isn't overkill is because it solves some difficult problems that seem to be fatal for LI and CHUNKS.

4. Against LI

The distinctive claim of lexical item theories is:

LI: Idiomatic phrases are semantically unstructured lexical items.

This is an attractive thing to say, because it explains both of the puzzling features of idioms. It explains UNPREDICTABILITY, because the apparent constituents of idiomatic phrases are *merely* apparent—the idiomatic phrases don't get their meanings from the meanings of their parts. And it explains INFLEXIBILITY for the same reason—if idiomatic phrases are just lexical items, we shouldn't expect other phrases with the same literal content to also be instances of the same lexical item.

So LI does a good job explaining the behavior of idioms that we've seen so far. My task in this section will be to show that, unfortunately, idioms also behave in ways that make it clear that LI can't be right. More carefully: I'll show that LI can't be the whole story, since at least some paradigmatic idioms go in for a kind of behavior that LI can't account for. (The phenomena I'll cite in this section will be familiar from the literature on idioms in linguistics. Most of the examples in this section are from Nunberg et al. (1994), or are modifications of their examples.)

One source of trouble for LI is that idioms aren't *completely* inflexible. Since LI seems to predict that idioms *will* be completely inflexible, this is a strike against LI. Let's have a look at some examples of flexibility:

Felicitous Passive
4.   a. The ice was broken by Paulie.
            b. Strings were pulled.
            c. Tabs were kept on Jackie by both the FBI and the NYPD.
            d. No stone was left unturned.

This is surprising if LI is true, but maybe it's not yet fatal. Maybe LI can allow, for example, that there are several entries (including one for the passive) for some idioms. But there are other ways in which we can modify idioms which seem to be incompatible with the claim that the parts of idioms are semantically inert. For example:

Modification of parts
5.   a. They left no legal stone unturned.
      b. We must beat our terrifying swords into plowshares.
      c. She kicked the filthy habit.
      d. Your remark touched a nerve that I didn't even know existed.
      e. Many Californians jumped on the bandwagon that Perot had set in motion.

Quantification
6.   a. That letter touched a couple of nerves.
      b. If this doesn't work, there are still a few more strings we could pull.
      c. How about if we just beat *most* of the swords into plowshares?

Topicalization
7.   a. Those strings, he wouldn't pull for you.
      b. His closet, you might find skeletons in.

Ellipsis
8.   a. Junior's goose is cooked, but Tony's isn't.
      b. We thought the bottom would fall out of the numbers racket, but it
            didn't.

Anaphora
9.   a. We thought tabs were being kept on us, but they weren't.
      b. I had a bone to pick with them, but they were so nice that I forgot
            about it.
      c. His family pulled some strings on his behalf, but they weren't enough
            to get him acquitted.

If idiomatic phrases were lexical items, we wouldn't be able to change their meanings by modifying their parts, because their parts would be semantically inert. But we *can* change the meanings of idiomatic phrases by modifying their parts. So their parts must be semantically ert. (And we can't explain what's going on as the modifiers modifying the whole phrase, rather than the parts—at least not without a lot of fancy footwork. The modifiers that we're using aren't of the right type to act on the sorts of things that are candidate semantic values of the whole phrase.)[7]

If LI were true, we wouldn't be able to quantify in to idiomatic phrases—the idiom wouldn't have any constituent noun phrases for the quantifier to hook up to. But we can quantify in to idiomatic phrases, so they must have semantic structure; in particular, some of them must have NPs as constituents.

If LI were true, you wouldn't expect to find topicalization—to topicalize something, there's got to be something there to topicalize. And if we can highlight sometimes one, sometimes another of the constituents of an idiom, the idiom had better *have* some constituents.

If LI were true, idioms wouldn't support ellipsis or anaphora. To have ellipsis within an idiom, you've got to have a VP. To have anaphora, you've got to have an NP. Since idioms do support ellipsis and anaphora, they must (some of them) have VPs and/or NPs as constituents. But that's incompatible with their being unstructured, and so it's incompatible with LI.

In short: LI predicts that idioms will not be susceptible to any sort of manipulations or modifications that require them to have semantically significant constituents. Since (many) idioms *are* susceptible to such manipulations and modifications, LI is not the correct theory of (those) idioms.

5. Against CHUNKS

Recall the key claim of chunky theories of idiom:

CHUNKS: The parts of idiomatic phrases have nonstandard semantic values when they occur within the idiom, which then get composed according to the usual compositional rules.

This view accommodates the evidence that made trouble for LI in the previous section, because it doesn't require that idiomatic phrases be semantically unstructured.

Here is an example of how CHUNKS breaks down an idiom into semantically active constituents (from Nunberg et. al (1994:497)): *spill the beans*, when it's used idiomatically, means *divulge the information*. According to CHUNKS, this is because, in idiomatic contexts, *spill* means *divulge* and *beans* means *information*. That's why we can modify the core idiom in the ways illustrated in the last section:

10. a. Tony thought Chris had spilled the beans, but it was actually Livia
   who spilled them.
  b. Jimmy spilled the incriminating beans.
  c. Jimmy spilled the beans that put Junior in prison.
  d. If Paulie hadn't whacked him when he did, Jimmy would have spilled
   all the beans.
  e. These beans, you won't spill if you know what's good for you.
  f. Silvio thought that the beans had been spilled, but they hadn't.

We can do all of this, according to CHUNKS, because *spill the beans* breaks up into

chunks (*spill* and *the beans*) to which we can assign reasonable semantic values (*divulge* and

the information) such that, when we compose them, we get a reasonable semantic value for the

idiomatic phrase (*divulge the information*), and that deliver the right truth conditions for

sentences in which the idiomatic phrase occurs.

But not every idiom breaks nicely into chunks. Take, for example, *kick the bucket* and

saw logs. There is no plausible way to assign alternative semantic values to *kick* and *the*

bucket such that when you compose them, you get *die* as a semantic value for the whole phrase.

Similarly, we can't give plausible alternative interpretations of *saw* and *logs* such that *saw logs*

turns out to mean *sleep* (or *snore*). For these sorts of idioms, it looks like the best we can do is

what LI told us we should do for all idioms: just give an alternative interpretation to the whole

phrase, and take it to be a semantically unstructured primitive. So *kick the bucket* just gets

interpreted as meaning *die*, and *saw logs* just means *sleep* (or *snore*).

Nunberg et al. call those idioms whose meanings are decomposable to their parts

idiomatic combining expressions, and those that resist such decomposition, and so need to be

treated as unstructured lexical items, *phrasal idioms*. I will adopt their terminology here.[8] A theorist who subscribes to CHUNKS will almost certainly believe (with Nunberg et al.) that idioms divide into these two classes, and that the LI view correctly describes the behavior of phrasal idioms, while CHUNKS correctly describes the behavior of idiomatic combining expressions. Such a theorist will predict that we'll see very different sorts of behavior from the two classes of idioms – in particular, they will predict that we won't see the sorts of felicitous modification that made trouble for LI as a general view of idioms in the case of phrasal idioms.

Now that we have CHUNKS in place, let's look at some phenomena – ones that, as far as I am aware, have not yet received much attention in the literature on idioms - that suggest that it's not right.[9] More carefully: they suggest that CHUNKS is not the whole story—we need more resources than CHUNKS allows us in order to account for some of the distinctive behavior of some prototypical idioms.

The first of these troublesome phenomena is a variation of the sort of felicitous modification of parts that made trouble for LI:

Figurative Modification
11.   a. The strings we've been pulling to keep you out of prison are fraying badly.
        b. That horse you're flogging isn't quite dead yet, but it's definitely not well.
        c. I know the bathwater was really dirty, but you still shouldn't have thrown the baby out with it.

We get the wrong contents for all of these if we apply the usual semantic values of *fraying*, *not well*, or *dirty* to the semantic values the advocate of CHUNKS is going to assign to *strings*, *horse*, and *bathwater*. For example, if *strings* means something like, *channels of influence* (and *pull* means something like *manipulate*), we'll get either the wrong truth conditions for the sentence or none at all if we use the usual semantic value of *fraying* when we give the semantics of the whole sentence.

So in these cases, we can't get the right truth conditions for the sentence by just substituting alternative semantic values for the idiom chunks.   We'll also need to provide alternative semantic values for modifiers—*fraying*, *not well*, and *dirty*—that don't occur in the core idiom. But this isn't an isolated phenomenon. Figurative modification happens a *lot*. So we won't just need to introduce a *little* bit of ambiguity, with different candidate semantic values for a few phrases that occur as idiom chunks. We'll need to introduce a *lot* of ambiguity, with different semantic values for every expression that can be used for figurative modification of an idiom chunk. This is probably going to be more or less everything.

The other troublemaking phenomenon is also best illustrated by example:[10]

Extendibility
12. *pull a rabbit out of a hat*[11]
     Brazil is a team with more rabbits in its hat than most.

13. *let the cat out of the bag*
     a. Junior let the cat out of the bag, but it was going to get out eventually anyway;
        The Feds had been sawing away at the burlap for weeks.

b. If you let this cat out of the bag, a lot of people are going to get scratched.

14. *get one's ducks in a row*
  a. I thought I had all my ducks in a row, but it must have been some kind of optical illusion.
  b. I had my ducks in a row for about a week, but then they just went flapping and squawking all over the park.
  c. Other people seem to be able to get their ducks in a row, but I think I've just got a really uncooperative bunch of ducks.
  d. Dude, what you need is a duck wrangler.

It's a bit hard to give a precise characterization of what's happening in all of these cases until we've got a theory in hand. (I'll offer a PRETENSE-based explanation in section 6.) But at least two things are clear: (a) the idiom extensions in (12)-(14) are perfectly felicitous, and (b) it's difficult to explain why they're felicitous, and how they come to mean what they do, if CHUNKS is the whole story about idiom interpretation.

As in the cases of figurative modification in (11), in order for CHUNKS to assign the right contents to the examples in (12)-(14), we're going to need alternative, idiomatic semantic values for a lot of words, phrases, and sentences that don't occur in the core idiom. In fact, it seems as if an *enormous* part of the language will have to be many ways ambiguous, since very many idioms are extendable in very many ways.

What's worse, it's not just idiomatic combining expressions that are extendable:

Extendable phrasal idioms
 15. a. Tony: Has he bought the farm yet?

Silvio: Nope. The offer's been accepted and the loan's been approved, but
he's taking his time closing on it.
b. Paulie: I hear Mr. Jones kicked the bucket.
Vito: Yeah. He almost connected yesterday; today he really put the boot on it.
c. Whoa, man, you need to take a couple more chill pills.
d. If you keep your fingers to yourself, they won't get burned.

The fact that even phrasal idioms are extendable is especially bad for CHUNKS. To accommodate extendibility in idiomatic combining expressions, the advocate of CHUNKS needs to say that many more expressions need to get special idiomatic interpretations than we at first suspected. This is pretty bad, because it introduces really a lot of ambiguity—as we noted before, *very* many expressions are going to be many ways ambiguous.

But the situation with extensions of phrasal idioms is worse. In particular, it's very bad for CHUNKS that sentences like 15.c, in which we're applying a quantifier to *chill pill*, which ought to be semantically inert if CHUNKS is true, are okay. It's also very bad that sentences like 15.a, 15.b and 15.d, in which we have pronouns referring back to *the farm*, *bucket* and *fingers*—which also ought to be semantically inert if CHUNKS is true—are okay. On a chunky theory, only idiomatic combining expressions ought to support the sort of quantification, anaphora, and modification of parts that made trouble for LI. Phrasal idioms still ought to work in just the way that LI said *all* idioms work. The trouble is that this doesn't seem to be right – these phenomena aren't restricted to idiomatic combining expressions in the way that the advocate of CHUNKS will predict that they should be.

I have offered two arguments against CHUNKS based on extendibility phenomena. The first is that assigning the right truth conditions to extensions of idioms is going to require assigning alternative semantic values to very many more expressions than just the ones that occur in core uses of idioms, and many expressions will need to be many ways ambiguous. This is a lot of ambiguity. Pretty much every expression can be used in a way that would, if CHUNKS were correct, call out for the assignment of a nonstandard semantic value in extending *some* idiom, and very many—probably most—can be used to extend many idioms, with different idiomatic semantic values required for the different extensions.

This is a bad consequence. It would be much better if we could account for our ability to extend, and understand extensions of, idioms without cluttering the lexicon with so much ambiguity, and the head with so many rules for when to resolve the ambiguity which way. It would be better if all we needed to have in our heads was the ordinary, literal meanings of the words, and some procedure for generating idiomatic interpretations.

The second based-based argument I offered against CHUNKS is that, even if we grant all of this ambiguity, it can't account for the extendibility of phrasal idioms.

A final, weaker criticism of CHUNKS (which also applies to LI) is that it leaves an important gap in our explanation of idiomatic meaning. Suppose we're happy to grant the many-ways ambiguity of most expressions. As Nunberg et. al. themselves point out, there are still non-accidental connections between the literal meanings of expressions and their idiomatic meanings.[12] These connections between literal and idiomatic meaning go unexplained if our

only story is that it's the idiom chunks, not the whole phrases, that are extra lexical items. But these non-accidental connections seem to be *essential* to our ability to extend, and understand extensions of, idioms. An account of our ability to generate and understand sentences involving idioms and extensions thereof that doesn't make reference to the connections between literal content and idiomatic content has left out something important.

6. Why a Pretense Account Would Help

Here is the central claim of a pretense theory again:

PRETENSE: The parts of sentences containing idioms all retain their usual semantic values, and are composed in the usual way, but the sentence is assigned nonstandard truth-conditions by processing its literal content through a pretense.

It's clear that this avoids the need to postulate very large amounts of ambiguity that we encountered with CHUNKS. We just need the usual, literal semantic values of the expressions. On the other hand, though we don't need lots of extra lexical items, we *do* need an extra step in the process of interpreting idioms—the step of processing the literal content through the pretense. This is a theoretical cost, but it's one that's worth paying. The first benefit is, as we've just seen, that we get to avoid all of the unwanted ambiguity that came along with CHUNKS. Another is that a pretense account allows us to give a nice explanation of the

extendibility of idioms. The short version of the story is that pretenses are extendable in the same way as idioms, so according to PRETENSE, idioms are extendable because, and to the extent that, their governing pretenses are extendable. The long version of the story follows.

It's pretty obvious that pretenses are extendable. Let's introduce a very simple pretense—the sole principle of generation is, *if someone's holding their hand* so*, then it's fictional that they have a gun in that hand*. (You know how to fill in for 'so', or maybe you know a couple of candidates. Pick your favorite to fill in the principle.) Go ahead and hold your hand in the right way to make it fictional that you're holding a gun. Now throw the gun to your other hand. Shoot some piece of furniture. Blow the smoke off of the barrel. Spin the gun around your finger. Put the gun back in its holster. I predict that you had no difficulty at all knowing what to do with your body in order to make it fictional that you were throwing, catching, firing, blowing smoke off the barrel of, spinning, and holstering your gun.[13] But we hadn't introduced any principles of generation of the form, *if somebody's φing, then it's fictional that they're (throwing/catching/firing/blowing smoke off the barrel of/holstering) a gun.* You had to introduce them 'on the fly' in response to my demands that you make some particular proposition fictional. You had to *extend the pretense* by introducing new principles of generation. And, if you're like most people, you did this effortlessly, without any conscious thought. So (a) pretenses are extendable, and (b) we're very good at figuring out how to extend pretenses 'on the fly', often without even noticing that we're doing it.[14]

It's easy to see how the extendibility of pretenses would account for the extendibility of idioms if PRETENSE is true. If we're having a pretend gunfight, we might start off with the simple pretense defined by the single principle of generation, *if somebody's holding their hand so*, *then they're holding a gun in that hand*. Then somebody tells us that Juan just shot Jim, blew the smoke off the barrel of his gun, then spun it around his finger and holstered it. We know right away how to extend the pretense in order to figure out what Juan *actually* did. So we don't need to have the principles of generation about shooting, spinning, blowing smoke, and holstering already in place in order to tell people things about Juan's actual behavior by telling them things about what's true according to the gunfight game. We can count on our audience to extend the pretense in the right way when we say something within the pretense, just as I could count on my audience (you) to extend the pretense in the right way when I told you what to do within the pretense.

If PRETENSE is right, the situation is just the same with extensions of idioms. We start off with the simple pretense that governs *kick the bucket*: *if somebody dies, pretend that there's some salient bucket that they kicked*. Then somebody says "Livia didn't quite kick the bucket, but she took a good strong swing at it". We know right away how to extend the pretense in order to figure out what has to have *actually* happened in order for it to be fictional that Livia took a good strong swing at the bucket, but failed to kick it.

This won't always be true—there are some extensions of idioms that don't work, either because they're very strained or because they're just uninterpretable. (E.g., "Livia kicked the

green bucket.") The pretense account predicts this. The same thing happens with explicit make-believe games. (Reporting on a game of mud pies, I say, "and then we had the lobster thermidor".) If a pretense theory is right, then what goes wrong in both of these sorts of cases is that we can't see any natural way to extend the pretense that would tell us what sorts of real-world facts would make *that* fictional. This shows us another potential benefit of a pretense account – it promises to help us to explain the difference between acceptable and unacceptable, interpretable and uninterpretable, extensions of idioms, in terms of the ease or difficulty of extending the underlying pretense in ways that would allow us to figure out the sorts of real-world conditions that would make that sort of thing fictional.[15]

This also looks like a plausible explanation of the uninterpretable idiom extensions. While it's pretty clear how to extend the bucket-kicking pretense in a way that will tell us what the world would have to be like for Livia to have taken a swing at the bucket without actually connecting, it's not clear how to extend the bucket-kicking pretense in a way that would tell us what the world would have to be like for it to be fictional that Livia kicked the *green* bucket.

So we have a nice explanation of extendibility if PRETENSE is true. Notice that our explanation works both for idioms that a proponent of CHUNKS would classify as idiomatic combining expressions and for those that the chunky theorist would classify as phrasal idioms: Since, according to PRETENSE, the sentences retain their usual constituent structure, there's no problem about modifying the parts of idioms like *kick the bucket*. We can figure out the

real-world implications of a wide variety of claims about what's fictional, provided that we know, or can figure out, the relevant principles of generation.

We also have, as we just saw, an explanation of the *limits* of extendibility. On a pretense account, extensions *should* fail when there's no way (or no clear best way) to extend the pretense—by introducing either new fictional facts or new principles of generation—in order to deliver reasonable real-world truth conditions for the utterance.

Finally, PRETENSE explains the systematic connections between the literal meanings of expressions that occur in idioms and idiom extensions and their idiomatic meanings.[16]

The only fly in the ointment is that, as Stanley (2002) points out, pretense theories overgenerate.[17] PRETENSE says that to interpret an idiomatic sentence, you first determine its literal content in the usual way, and then process it through the pretense. So we ought to get the same truth conditions for any sentence with the same literal content. But we don't. A lot of sentences with the same literal content don't get idiomatic interpretations at all. Or at least, their idiomatic interpretations are strained and not terribly natural. Remember the examples of inflexibility from the beginning of the paper:

1. (passive)
    a. Tony blew off steam.
    b. *Steam was blown off by Tony.
2. (anaphora)
    a. Tony shot the breeze with Junior, and Paulie shot the breeze with Silvio.
    b. *Tony shot the breeze with Junior, and Paulie shot it with Silvio.
3. (substitution of synonyms)

      a. Richie kicked the bucket.
      b. *Richie kicked the pail.

There's definitely something wrong with the starred sentences, but (it seems) as far as PRETENSE is concerned, they ought to be just fine.[18] This is a really big problem—probably fatal if we can't find a way to fix it. INFLEXIBILITY is one of the signature marks of idioms, and it does seem that an account of idiom that doesn't explain INFLEXIBILITY fails, as Stanley says, "as badly as it is possible for an account of idiom to fail." So however well PRETENSE does at explaining extendibility and the connections between literal and idiomatic meaning, we'll have to send it back to the scrap heap if we can't square it with INFLEXIBILITY.

7. Solving the Inflexibility Problem

Idiomatic phrases, and especially modified or extended idiomatic phrases, like the ones we've been discussing, aren't *always* used idiomatically—they can also be used to express their *literal* contents. Sometimes people really do kick buckets, saw logs, spill beans, and let cats out of bags, and we can tell each other about these goings-on without having to resort to lots of fancy circumlocution. So the pretense theorist needs some story about how we know when to 'switch on' a given pretense—when to bring the relevant principles of generation into effect and start processing sentences through them. I think that looking at the beginnings of how this story would go can show us how to solve the overgeneration problem.

Notice that familiar kinds of explicit pretenses aren't always in effect, either. The pretenses that govern cops and robbers and mud pies are also active at some times and not active at others. Sometimes when I say, "there are four bank robbers behind you", the thing to do is interpret my utterance through the cops-and-robbers pretense and act accordingly—perhaps by spinning around, pointing your index fingers at the people behind you and shouting "bang! bang! bang! bang!". Sometimes when I say, "there are four bank robbers behind you", the thing to do is to interpret my utterance literally, and to very slowly lie down on the ground and put your hands on your head. It would be bad to get confused about which kind of situation we were in.

So how do we know (a) when to interpret through a pretense and when not to, and (b) which pretense to interpret through when a pretense is called for?

There are two ways to initiate a pretense. The first is to just start playing the game and trust your partners to catch on. So, one way to bring a pretense into effect is to do something that only makes *sense* if that pretense is in effect. For example, if you swagger into a colleague's office with your hands near your belt and drawl, "Professor Smith, this department ain't big enough for the two of us!" you will (if Professor Smith cooperates) have started a wild west showdown game.

At least some of the examples of idiom and pretense extensions in the previous sections were also examples of this sort of phenomenon—we can bring a new principle of generation

into effect in an ongoing pretense by doing or saying something that only makes sense if that principle of generation is in effect.

This is much like what happens when conversation partners accommodate each other's presuppositions about the world or about the conversational context—cooperative conversation partners adjust the context so that what the other parties to the conversation are saying and doing makes sense, sometimes by adjusting their view about what the world is like, sometimes by changing the standards of precision that are in effect, and sometimes by initiating or extending a pretense.

This method for initiating and extending pretenses isn't always reliable, though. We have to rely on our partners to figure out what the relevant pretense is, so we have to give them enough information to figure out what to do in order to accommodate us. This is simple enough with the western showdown game. But there are a lot of cases where we can't reasonably expect our audience to figure out, just from what we've done or said, which pretense we're trying to introduce.

Complex or obscure games will be like this. I can't (or at least, I can't easily or reliably) start a game in which we're corrupt undercover Treasury agents in 1928 Des Moines infiltrating a counterfeiting ring in order to take it over so that we can mint enough money to finance the startup of our own bootlegging operation, just by starting to play and hoping that you catch on. (What's making the trouble here is partly the complexity of the game, and partly the obscurity of some of the things that I want you to pretend.) For these sorts of cases, we'll

need another method for letting our audience know which pretense to use in interpreting our utterance and/or behavior.

The other way to initiate a pretense is to give a cue. The simplest sort of cue is to just tell your partners what to pretend—either which game to start playing, or which principle of generation to add to an existing game. The standard childhood cues in my neighborhood were, "let's play…" (as in, "let's play cops and robbers") and "let's say…" (as in, "let's say that puddles are quicksand"). Cues needn't be so explicit as that, though. If we play cops and robbers a lot (better, if we play a particular, somewhat complicated variation on cops and robbers, like the Treasury agents game mentioned above), it might be useful to have a more conventionalized cue. In principle, the cue could be anything—a hand signal, a stock phrase, the last line of dialogue from the last time we played, etc.

The solution to the overgeneration problem, I think, is to say that a similar cuing phenomenon is at work in the interpretation of idioms. Use of the particular form of words serves as a cue to initiate the relevant pretense. We get inflexibility because, if you gratuitously alter the form of words, you're not providing your audience with the right cue.

We also get an explanation of why the starred sentences in (1)-(3), and manipulations and extensions of idioms in general, sound better if they follow uses of the idiom in its canonical form: Once the pretense has been activated by using the right form of words, it's likely to remain in effect for later utterances, so the cueing effect of the precise idiomatic phrase becomes less important.

But not *completely* unimportant. Deviations from the precise form of words *to no good purpose* will be less felicitous than deviations that provide some communicative benefit. The presupposition that I'm being cooperative will make the idiomatic readings of *idle* deviations from the canonical form of words more strained than the idiomatic readings of extensions that offer a communicative payoff by conveying something that the precise idiomatic phrase does not. Even once a pretense has been initiated, there's always a question, in interpreting later utterances, of whether or not it's still in effect – whether or not later utterances are to be interpreted literally, or through the relevant pretense. There's also a live question of *which* later utterances are to be interpreted through the pretense – after all, even after I say "Tyler spilled the beans", it's not as if we interpret *everything* that anybody says thereafter through the bean-spilling pretense. Even after a pretense has been initiated, there's a question, in interpreting later utterances, of (a) whether the pretense is still in effect or not, and (b) if it is in effect, whether or not the present utterance falls under its scope.

This is why, even after a use of the canonical form of a given idiom, *gratuitous* deviations from the canonical form still display the distinctive sort of deviance, defectiveness, or oddity that makes them deserve stars beside their sentence numbers. If we're being maximally helpful and cooperative conversational partners, we'll give our interlocutors as much assistance as we can in answering interpretive questions such as (a) and (b) above. So when trying to convey something that could easily be conveyed using the idiom's canonical form, maximally helpful conversational partners will use the canonical form. Given this fact,

failure to use the canonical form when the canonical form would have done just as well (e.g., using "boot the pail" rather than "kick the bucket") will, to the extent that one thinks that one's conversational partner is being maximally helpful and cooperative, generate some pressure *not* to interpret their utterance idiomatically, and will make the idiomatic interpretation awkward, strained, or odd-sounding. (Since, if the speaker meant their utterance to be interpreted idiomatically, they could have made your interpretive job easier, at no communicative cost, by using the canonical form.)

Obviously, failing to use the canonical form of words when the canonical form would have done just as well (for communicative purposes, at least) doesn't make the idiomatic interpretation completely unavailable. These sorts of gratuitous modifications happen all the time, and people track them remarkably well. That's because everybody knows that their conversational partners aren't *just* interested in being maximally interpretively helpful. They're also interested in comedy, originality, wit, etc., and all of these things can speak in favor of deviations from an idiom's canonical form, even when the canonical form would do just as well for information-transfer purposes.

Another potential benefit of a PRETENSE account is that it suggests a way to account for (at least some of) the variation in how flexible different idioms are, in terms of the relative importance of the particular form of words as a cue to engage in the appropriate pretense. One respect in which idioms differ from one another is in just how unpredictable their meanings are. On a PRETENSE account, this is because the relevant pretenses might be more

or less obscure for different idioms. (Compare *kick the bucket, shoot the breeze,* and *his goose is cooked* on the one hand and *pull strings*, *let the cat out of the bag*, and *beat swords into plowshares* on the other.)

Call an expression *weakly unpredictable* if its meaning isn't the one that's generated compositionally from the constituent terms' conventional meanings. Call an expression *strongly unpredictable* if there's no reasonable way in which someone previously unacquainted with the expression, but equipped with all of the conversational abilities required to interpret people's utterances generally (including metaphorical ones), could figure out the idiomatic meaning on her own.[19]

While all idioms are (at least) weakly unpredictable, they're not all strongly unpredictable. Figuring out from scratch what's meant by "pull strings", "spill the beans", or "let the cat out of the bag" isn't a hopeless task in the way that figuring out what's meant by "kick the bucket" or "shoot the breeze" is. In the case of idioms that are only weakly unpredictable, there's some interesting, fairly natural relationship between the kinds of situations described by the idiomatic phrase in both its literal and idiomatic interpretations, in virtue of which the idiom seems natural or appropriate.[20]

It's an interesting fact that the most flexible and extendable idioms seem to be those that are only weakly unpredictable. These are the idioms that most easily, and most happily, support grammatical transformations, substitution of synonyms, etc., and that are easiest to extend. "Let the cat out of the sack" and "strings were pulled" are fine, while "kicked the pail"

and "the breeze was shot" are awful. (The differences in flexibility are particularly pronounced when we think about the acceptability of these sentences as *first* uses of the relevant idioms.)

Here is a possible explanation of this fact: the weakly unpredictable idioms allow more flexibility because the verbal cue that tells the hearer which pretense to engage in is less important—it's still pretty easy to figure out what the relevant pretense is, even without the cue of the particular idiomatic form of words. Strongly unpredictable idioms allow less flexibility (especially in first uses) because the verbal cue is much more important. The relevant pretense isn't one that we can easily figure out on the fly, so if we want our audience to interpret the idiom correctly, it's important that we give them the right conventional cue that tells them which pretense to engage in.

The difference in the importance of the verbal cue will be a matter of degree. Even weakly unpredictable idioms aren't *completely* flexible. Given a pretense view, this shouldn't be surprising. For one thing, as noted above, it's important to have a cue not just to engage in the *particular* pretense that governs the idiom, but also to have a cue to engage in pretense *at all*, rather than just interpreting the utterance completely literally. Even for the most transparent of idioms, it's important to signal to one's conversational partners that they should interpret figuratively rather than literally. And while there are various pragmatic ways of doing this, the easiest way is to use the right form of words. Failure to use the right form of words, then, will generate some resistance to interpreting the utterance idiomatically, since there's a presumption of cooperativeness, and a cooperative speaker will use the easy way of signaling

idiomaticity unless there is some payoff (communicative, comic, or whatever) for doing otherwise.

8. An Objection

I've leaned very heavily on the extendibility of idioms in arguing against CHUNKS, and in favor of PRETENSE. But perhaps there's something funny about extensions. Maybe there's one story to tell about the interpretation of core uses of idioms (say, what CHUNKS says is going on), and another story to tell about the interpretation of extensions (maybe even what PRETENSE says). There does seem to be a felt difference between the two kinds of cases, and perhaps this felt difference should lead us to adopt a two-pronged theory.[21]

This seems like a reasonable thing to be worried about, and a defender of a unified PRETENSE view ought to have something to say in response.

The hypothesis that there's a single interpretive process at work in both core uses and extensions, and the hypothesis that there's one process for core uses and a different one for extensions, will certainly wind up making some different predictions, which we could recruit some subjects and test out. Though I've obviously got some suspicions about what the right story is here, what we'll want to say about this at the end of the day is going to depend on empirical results.

For example:[22] Ambiguous expressions display a distinctive sort of priming behavior. When speakers hear a sentence containing an ambiguous expression, there is a very brief

period during which all of the candidate meanings are activated, and the speaker will be quicker to recognize words that are semantically related to *any* of the candidate meanings as meaningful words. This effect fades very quickly—after about 750 milliseconds, subjects are still quicker to recognize words related to the contextually relevant meaning of the ambiguous term, but the speed with which they recognize words related to the *other* meanings returns to baseline.[23] This leads to two predictions of any theory of idiom according to which expressions that occur in idioms have more than one candidate meaning (like CHUNKS, for example):

First, if the expressions that occur in idiomatic combining expressions really are ambiguous—if, for example, "spill" sometimes means *divulge*—then we should expect to see priming effects for words semantically related to the *idiomatic* meanings of the components of idiomatic combining expressions, even when they occur in clearly *non-idiomatic* contexts. For example, we should see priming of words semantically related to *divulge* very shortly after clearly non-idiomatic occurrences of "spill".

Second, in clearly *idiomatic* contexts, the priming effects on words related to the *literal* meanings of the expressions should be suppressed very quickly. Within 750 milliseconds or so after an idiomatic occurrence of "kick the bucket", we shouldn't see any priming for, for example, "pail".

As far as I know, no one has done the relevant experiments, and so I don't know which way the results would go. If I had to guess, I'd guess that the ambiguity theory's predictions

wouldn't be borne out. My suspicion is that hearing non-idiomatic occurrences of "spill" or "cat" does not prime one for semantic relatives of "divulge" or "secret", and that hearing idiomatic occurrences of "kick the bucket" *does* prime one for such semantic relatives of "bucket" as "pail". But, that's just my suspicion. There's a long history of surprising experimental results, and what will ultimately decide these issues is which way the results actually go, not which way it seems plausible to think that they'll go before we run the tests.

If and when we do this sort of investigation, we might find that we should say different things about different idioms. Perhaps a two-process account – where the core uses of the idiom in its canonical form are interpreted in the way that CHUNKS or LI describe, but the recalcitrant extensions trigger a very different interpretive process, and these are processed in the way PRETENSE describes – is right for some idioms, while a PRETENSE-only account is right for others. I don't think we should be surprised if this turns out to be the case. Here is an attractive (if not very original) picture of the life cycle of (certain kinds of) idioms that would predict this result:

Some idioms begin life as novel metaphors. Let's suppose that some metaphors have a determinate real content, which they get via a pretense-based process. Early on, the metaphor is interpreted from scratch, on the fly, each time that it's encountered. But then some metaphorical phrase (call it M) catches on, and starts to be used very often, and always with the same intended meaning. (And so always with the same associated pretense—call it P).

After a while, if M is used often enough, people start to cotton on to the fact that every time that string of words is used non-literally, the right pretense to interpret it through is always P. Eventually, encountering M starts to serve as a cue to interpret through P, and an idiom (one that behaves in the way PRETENSE describes) is born. M will be weakly unpredictable if it's fairly easy to figure out on the fly that P is the right pretense to interpret it through, strongly unpredictable if it's very difficult to figure out. It might happen that, over time, the community changes in ways that make P harder to recover on the fly, and a weakly unpredictable idiom will become strongly unpredictable.

It might also happen (especially if M is strongly unpredictable) that for a long time, M is rarely or never extended or used in a variant form. (Maybe there's not much representational benefit to extending or modifying, or maybe users just don't happen to do it.) After a while, people might start taking a shortcut in interpreting M: rather than always interpreting it through P, just treat M (or maybe some of M's parts) as meaning something other than their literal meaning. Now we'll have an idiom that works in the way that CHUNKS or LI describe. Suppose that this happens. Creative speakers may still be able to "reactivate" M by passivising, topicalizing, quantifying in, or extending. When this happens, interpreters will reactivate P, and interpret this new M-extension through the pretense (or an extension thereof), even though core uses of M are no longer being interpreted through P.

(Another possibility is that the idiom is reactivated as a metaphor, but one that's interpreted through some pretense other than P. This would be likely to happen in cases where

the original metaphor exploited some cultural or linguistic fact that's changed since the phrase came into popular use – present-day modifications of, e.g., *hoist on ___'s own petard* and *the whole nine yards* seem likely to be like this.)

Conceivably, it could eventually happen that M freezes completely, and no one, no matter how clever or creative, will be able to produce felicitous alterations or extensions of M, or to reactivate P. Then I think we will have seen the death of an idiom and the birth of a regular old boring lexical item, which won't display any of the sorts of extendibility behavior discussed here.

If some idioms have this sort of career, then the right story about idioms will be heterogeneous. Some idioms will always be interpreted through a pretense, while others will be interpreted chunkily or as simple lexical items in their core uses and through a pretense in extensions. The line between idioms and metaphors will probably be quite vague. I don't want to take a definitive stand on whether any idioms *do* have this sort of career, but I suspect that some probably do. This is one reason why I don't want to insist too much on PRETENSE being the *unique* right theory of idiom.

Conclusion

Neither LI nor CHUNKS can account for the extent to which many idioms are flexible and extendable. PRETENSE can, and the overgeneration worry can be defused. So we should

endorse a pretense theory of idiom interpretation, in order to account for the full range of our abilities to produce and interpret idiomatic sentences.

It's worth pointing out that I haven't given any arguments that there *couldn't* be expressions that work the way LI and CHUNKS say idioms do. All I've set out to show is that many paradigmatic idioms like "pull strings", "spill the beans", and "kick the bucket" don't work that way. So I don't really want to claim that, every time we get UNPREDICTABILITY, it's because we've got PRETENSE. The minimal claims that I hope to have established are that (a) very many paradigm idioms behave in ways that LI and CHUNKS can't accommodate, (b) this kind of behavior is nicely explained by PRETENSE, and (c) despite initial appearance, INFLEXIBILITY doesn't make PRETENSE a non-starter. If I've succeeded in this, PRETENSE should look much more attractive now than it did at the beginning of the paper.

References:

Crimmins, Mark (1998) Hesperus and Phosphorus: Sense, Pretense, and Reference, *Philosophical Review* **107**: 1-47

Davies, Martin (1983) Idiom and Metaphor, *Proceedings of the Aristotelian Society*. **83**: 67-86

Glasbey, Sheila (2003). Let's Paint the Town Red for a Few Hours: Composition of Aspect in Idioms, in *Proceedings of the ACL workshop The Lexicon and Figurative Language*, Wallington, ed., ACL 2003, Sapporo, Japan.

Katz, Jerrold (1973). Compositionality, Idiomaticity, and Lexical Substitution, in *A Festschrift for Morris Halle*, ed. Stephen R. Anderson and Paul Kiparsky, 357–76. New York: Holt, Rinehart, and Winston.

Kroon, Frederick (1992) Was Meinong Only Pretending?, *Philosophy and Phenomenological Research* **52**: 499-527

Marantz, Alec (1997) No Escape from Syntax: Don't Try Morphological Analysis in the Privacy of your own Lexicon, in *University of Pennsylvania Working Papers in Linguistics 4:2: Proceedings of the 21st Annual Penn Linguistics Colloquium*, 201-225.

McGinnis, Martha (2002) On the Systematic Aspect of Idioms, *Linguistic Inquiry* **33**: 665-672.

Moran, Richard (1999) Metaphor, in *A Companion to the Philosophy of Language*, Hale and Wright, eds. (Blackwell)

Nunberg, Geoffrey, Sag, Ivan and Wasow, Thomas (1994) Idiom, *Language* **70**: 491-538.

Stanley, Jason (2002) Hermeneutic Fictionalism, in *Figurative Language*, French and Wettstein, eds. (Blackwell)

Swinney, David (1979) Lexical Access During Sentence Comprehension: (Re)Consideration of Context Effects, *Journal of Verbal Learning and Verbal Behavior* **18**: 645-659.

Swinney, David and Cutler, Anne (1979) The Access and Processing of Idiomatic Expressions, *Journal of Verbal Learning and Verbal Behavior* **18**: 523-534.

Walton, Kendall (1990) *Mimesis as Make Believe*, Cambridge, Harvard University Press

Walton, Kendall (1993) Metaphor and Prop Oriented Make-Believe, *European Journal of Philosophy*, **1**:39-57

Walton, Kendall (2000) Existence as Metaphor?, in *Empty Names, Fiction and the Puzzles of Non-Existence*, Everett and Hofweber, eds., Stanford, CSLI Press.

Yablo, Stephen (1998) Does Ontology Rest on a Mistake?, *Proceedings of the Aristotelian Society*, Supp. Volume **72**: 229-261.

Yablo, Stephen (2000a) A Paradox of Existence, in *Empty Names, Fiction and the Puzzle of Existence* Hofweber, ed. (CSLI).

Yablo, Stephen (2000b) A Priority and Existence, in *New Essays on the A Priori*, Boghossian and Peacocke, eds. (Oxford).

Yablo, Stephen (2002) Go Figure, in *Figurative Language*, French and Wettstein, eds. (Blackwell)

[2] Stanley (2002)

[3] The account of pretense that follows is due to Kendall Walton (1990; 1993), and employed by, in particular, Stephen Yablo (1998; 2000a; 2000b; 2002) and Mark Crimmins (1998).

[4] The "or within"s here are in order to head off a possible problem about names, predicates, and sentences that fail to pick out any individual, property, or proposition. There's a problem, for example , about the semantic value of "Cowboy Bob" in, "Cowboy Bob just shot a buffalo". Since there's no such person as Cowboy Bob, we shouldn't say that "Cowboy Bob" has its usual semantic value, which makes its usual contribution to the truth conditions of the sentence, which tells us what to pretend (tells us something about the pretense), and then we use the principles of generation to work back to the real content. We can finesse this problem by talking about pretend-assertions (assertions *within* a pretense) rather than assertions *about* a pretense; within the pretense, "Cowboy Bob" *does* refer.

The empty names problem is a problem about missing objects. There are analogous problems about missing properties and relations, and about missing propositions. (Walton (2000) has a helpful discussion of these, as well as the empty names problem.) It's not clear that there really is a property or relation available for, e.g., *make no bones about*, *steal ___'s thunder*, or *go haywire* to express, and if there aren't properties or relations for the predicates to literally express, there won't be propositions for the sentences to literally express. Walton's (2000) solution to here is the same: the pretense theorist should say that what's supposed to be fictionally true is not *the literal content of the idiomatic sentence*, but *that the idiomatic sentence expresses a true proposition*. So we don't need there to actually be any properties or propositions for the predicates or sentences to literally express – all we need is that, within the pretense, there are properties and propositions that they express.

Another option, of course, is to say that at least some of these idioms, where the sentences don't seem to express any proposition that one could take to be fictionally true, fall outside of the scope of the pretense theory – something like LI or CHUNKS is true of them. This needn't be such a horrible concession – as noted in the introduction, it's pretty unlikely that all idioms fall under a single natural kind, and so it's not a disaster if one winds up giving different theories of different classes of idioms, provided that there are some grounds for thinking that they really do work differently. My own view is that the pretense theorist ought to make both of these moves, in different cases, depending on whether the idiom in question displays the sort of behavior that motivates a pretense theory (in particular, the sort of susceptibility to creative extensions that I'll be making a big fuss about in section 5).

There are, obviously, going to be some complications here. But in the interests of space, I will (aside from returning to it a couple of times in footnotes) suppress this issue in the remainder of the paper. (Thanks to Jason Stanley, Jamie Tappenden and an anonymous referee for this journal for pressing me on these points.)

[5] Both Walton and Stephen Yablo also advocate a pretense theory of idioms. Yablo offers this sort of account of the interpretation of figurative language in general, and some of his examples are expressions that are most plausibly classified as idioms. Walton (2000) argues explicitly for a pretense theory of idioms.

[6] Another thing that I won't do here is look at competing theories of metaphor to see whether any of them can offer the same advantages as PRETENSE when applied to idioms. This would certainly be worth doing, but again, doing it here would take us too far afield.

[7] What about "ert", "to tract", etc.? "Inert" and "tractable" really *are* unstructured—"ert" and "tract" aren't meaningful constituents—but we can still understand sentences like "Their constituents are semantically ert after all", and "Bennett was the first to succeed in tracting this notoriously difficult problem". So why should we think that the sort of flexibility that we find in idioms is a sign that they were semantically structured all along, rather than taking flexings of idioms to be some sort of linguistic game-playing of the same kind as uses of "ert" and "to tract"?

Well, for one thing, the *feel* of game playing that's present in uses of "ert" and "to tract" isn't present in (at least very many) of the examples of idiom flexibility. But more importantly, what's going on in cases of idiom flexibility is much more complicated than what's going on with "ert" and "tract". We can recover the intended meanings of "ert" and "tract" pretty easily by means of some conjectural etymology that exploits the usual roles of prefixes and suffixes like *in-* and *–able*. If we know the usual role of *in-* and the meaning of "inert", it's easy to figure out what "ert" has to mean. The sorts of modifications that idioms allow are much more complex (especially the sorts of modifications we'll see in the next section), in ways that force us to tell a very different sort of story—they're not explainable in terms of the sort of conjectural etymology that allows us to explain "ert" and "to tract". (Thanks to Karen Bennett, Martin Davies, and Stephen Yablo for questions and discussion here.)

[8] Some examples of idiomatic combining expressions: pull strings, let the cat out of the bag, spill the beans, flog a dead horse, throw the baby out with the bathwater. Some examples of phrasal idioms: kick the bucket, saw logs, see red, buy the farm, burn one's fingers.

[9] McGinnis (2002) and Glasbey (2003) discuss some examples of the same type as the ones I discuss here, but use them to rather different ends. Walton (2000) invokes some examples of the same sort of phenomena I draw attention to here in the course of a brief argument for a pretense theory of idioms, though he does not go into detail about just how the competing theories of idioms founder on these phenomena, or the mechanisms by which a pretense theory can accommodate them.

[10] Actually, I think it's probably not another kind of phenomenon—figurative modification is just a type of extension. Thanks to William Lycan for helping me to get clear on this.

[11] This example is from the Australian announcer during the 2002 World Cup final.

[12] See for example Marantz (1997), McGinnis (2002), and Glasbey (2003) for discussions of some connections between the roles of expressions in literal and idiomatic contexts.

<hr>

[13]Actually, sometimes people – especially philosophers – have trouble with the spinning around the finger, because often they're already using the relevant finger as the barrel of the gun, and they don't want to break continuity. This is funny to watch. Thanks to Campbell Brown for pointing out this phenomenon.

[14] A possible worry about this example: maybe we weren't really extending on the fly, since most of us already knew that game. Maybe so. But notice two things: First, it would have been just as easy for you to do *unfamiliar* things with the gun: scratch your nose with the barrel, clean under your fingernails with the front sight, hammer a nail with the handle, etc. Second, other games. that you haven't already been exposed to, are just as easy. Here's the start-up principle of generation for a new game: if you're standing up just a little bit on your toes, so your heels are about an inch off the ground, pretend that you're wearing cowboy boots. Now, make it so you're wearing cowboy boots. Now sit down and take one of your boots off. Shake the rocks and dirt out of it. Spit on the toe and polish it up a little bit with your shirt. Put the boot back on and mosey around a little to make sure all the rocks are out of it. I'm guessing that that you had no difficulty with that. It is, in general, *remarkably* easy to make up new, unfamiliar games and to extend them almost indefinitely. (Thanks to Stephen Yablo for discussion of this point.)

[15] Note that there are actually two ways in which an extension could fail. The first way is the way it fails in the lobster thermidor case: we know perfectly well what we're being asked to pretend, but we're at a loss as to what sorts of real-world facts could make *that* fictional. Another possibility is that an extension could fail because we're at a loss as to what we're being asked to pretend. If I try to extend the game we're playing by saying, "let's pretend Fido is a bear", it's easy enough for you to go along, since you know what you're being asked to pretend. If I say, "let's pretend Fido is a tove", it's harder for you to go along – since you don't know what a tove is, you might well be at a loss as to what to pretend. (Though you might not – sometimes, this sort of extension works, by roughly the mechanisms discussed in footnote 9: In order to accommodate, it needn't be that "tove" actually picks out some kind K such that you pretend that Fido is a K. Instead, it could be that what's pretended is that "tove" picks out some kind that Fido is a member of – within the pretense, "tove" picks out some kind that, within the pretense, Fido belongs to. We'll wind up with a pretense that lacks any very determinate fictional facts about what Fido's like (though we could fill these in by further extensions, where we fill in some details about what sorts of things toves are), and this might undermine the worthwhileness of engaging in the pretense, which would give rise to some resistance to accommodation.) (Thanks to Jamie Tappenden for drawing my attention to this distinction.)

[16] Well, that's not quite right. It explains the role that the literal meanings of the constituent expressions play in determining the idiomatic truth conditions of sentences in which they occur. The expressions don't, according to PRETENSE, have any special idiomatic meanings.

[17] Stanley (2002) actually raises a number of objections to many different sorts of pretense theories. Here is not the place to respond in detail to the whole of that paper. What I'm about to do in the main text is respond to the principal objection – an extremely *good* objection – that he makes specifically against a pretense theory of idioms. But let me add just a few words in this note about the general concerns that he raises in the rest of the paper. Stanley has five general worries about pretense theories: 1) He argues that we ought not to believe fictionalist accounts of any area of discourse, since they will fail to be compositional – the meanings of the sentences won't be determined in a straightforward way be the meanings of their parts and the way in which those parts are put together. And it's true that pretense theories (such as the one I'm advocating for idioms) need more resources

than just the meanings of a sentence's component parts and the way in which they're assembled in order to deliver the idiomatic meanings that they want to assign to the sentences – they also need the principles of generation that govern the pretense that the sentence's literal content is to be processed through. But the motivation for compositionality – our capacity to generate and understand infinitely many novel sentences despite our finite supply of cognitive resources – doesn't require compositionality as such. It just requires some account of the understanding and generation of novel sentences such that our ability to do these things doesn't depend on some piece of magic. And while the precise details of how pretenses go in and out of effect, get extended, and so forth, are surely going to be complicated, they're not going to be *magic* – they're not going to require infinite amounts of storage capacity or processing power. We do in fact manage to keep track of pretenses all the time, so the extra ability that the pretense theorist needs to attribute to us is one that we certainly have, and so it's certainly one that only requires finite cognitive resources. 2) Stanley suggests that fictionalism is just the bad old method of paraphrase in sheep's clothing. It's certainly true that both fictionalist/pretense theories and paraphrase strategies are ways of trying to associate a sentence which, on its face, seems to have one bunch of truth conditions, with some other bunch of truth conditions instead. But pretense theorists, unlike simple paraphrase advocates, offer a (non-magical, compatible with the actual constraint that motivates compositionality) account of *how* the sentence gets hooked up with their favored truth-conditions. In this respect, fictionalists/pretense theorists appear to do better than advocates of a simple paraphrase strategy. 3) In many of the cases where fictionalist strategies are deployed for ontological purposes, Stanley rightly points out, it sure doesn't *seem* as if we're doing any pretending, or speaking less than perfectly literally. If we have reliable first-person access to whether we're pretending, and whether we're speaking literally, then fictionalist accounts of these domains of discourse will be dead in the water. A standard fictionalist response to this sort of objection is that we *don't* have such access. We can be engaging in a pretense (or doing something that's like engaging in a pretense in the relevant respects – we can represent some propositions in a less-than-completely-serious spirit, and maintain some rule-governed coordination between what we actually believe and what we represent in this spirit) without it being obvious to us that we are. And whether this response succeeds or not, the phenomenological facts are much less troubling for pretense theories of idioms than they are for, e.g., fictionalist theories of mathematics. In the case of idioms, it's typically clear to us that we *are* speaking figuratively, and the introspective data here are much less hostile to the idea that we are, in some less-than-completely-serious spirit, representing the literal content of the idiomatic sentences. 4) Stanley argues roughly as follows: Since the psychological capacity that enables us to engage in pretense is absent in people with autism, we should expect them to be unable to understand the bits of language about which a fictionalist/pretense theory is true. Since people with autism have no trouble understanding mathematics, etc., fictionalist theories of mathematics, etc. are false. Whatever its merits as an argument against pretense accounts of mathematics and so forth, this doesn't look promising as an argument against pretense theories of idioms, since people with autism do, in fact, have a notoriously difficult time with idioms (and with figurative language generally). 5) Stanley's fifth complaint is a general concern about the ontology-reducing motivation for fictionalist accounts of, e.g., mathematics. But as that motivation is not what's at work here, this objection won't apply to the case of idioms. Stanley is also concerned that there will be a general overgeneration problem for fictionalist/pretense theories across the board. In the main text, I'm going to address the sort of overgeneration problem that arises for pretense accounts of idioms. I'm not going to address the overgeneration problems that arise for pretense theories of anything else, or general questions about overgeneration across the board. Again, my interest here is just with the prospects for a pretense theory of idioms - I have no stake, for these purposes, in whether a pretense theory of anything else is right. (Except for metaphor, and I think the same extendibility phenomena provide pretty good motivations for pretense theories of metaphor, and Stanley's complaints mostly don't apply there.) Finally, Stanley also objects that there don't seem to be any distinctive features of idioms that a pretense theory helps to account for that a straight LI theory couldn't handle. I've

pointed to two types of such phenomena here: (1) the phenomena (such as felicitous quantification, anaphora, and modification of parts) that motivate CHUNKS, and (2) the extendibility phenomena that I argued CHUNKS can't accommodate.

[18]It's a good question exactly *what* is wrong with them, though. (A good question that was asked by both Karen Bennett and William Lycan in presentations of an earlier version of this paper—I am grateful to them, and to Elisabeth Camp, for helpful discussion.) We'll get an answer (or at least a partial one) out of the solution to the inflexibility problem in the next section. Here is a preview: what's wrong with the starred sentences is that, by gratuitously deviating from the standard form of words (rather than deviating from the standard form in order to meaningfully extend, etc.), you're failing to be a helpful and cooperative conversational partner.

[19] Thanks to Elisabeth Camp for extremely helpful discussion on this point, and for this formulation of the distinction.

[20] Though as Martin Davies (1983) points out, the idiomatic meaning's seeming natural once we know it can come apart from the realistic possibility of figuring it out on the fly. This suggests that there are some interesting distinctions to be made that are more fine-grained than the one I'm making between strongly unpredictable and weakly unpredictable idioms.

[21] There's also a felt difference between idioms and metaphors – as an anonymous referee for this journal pointed out, using a metaphor seems to *invite* extensions in a way that using an idiom doesn't (or doesn't necessarily). When you use a metaphor and I creatively extend it, I am pretty clearly just continuing the same sort of activity that you started. When you use an idiom and I creatively extend it, there's a bit of a feeling that I've somehow *changed the rules*. This suggests that perhaps there really is (a) something quite different going on in core uses of idioms than in extensions, and (b) something quite different going on in core uses of idioms than in metaphors. I definitely acknowledge the phenomenology the referee points to. I don't think, however, that it's present in *all* of the cases of extension that make trouble for LI and CHUNKS – figurative modification, for example, often seems to be of a kind with the core uses of the idioms. I'm also not sure how much theoretical weight to assign to this sort of phenomenology. This is, I think, a serious point that's deserving of more discussion than I can afford it here.

[22] The correct parts of the next few paragraphs are heavily indebted to Alec Marantz.

[23] Swinney (1979).