Semantics & Pragmatics Volume 19, Article 5, 2026
https://doi.org/10.3765/sp.19.5

This is an EARLY ACCESS version of

Grano, Thomas. 2026. Direction of fit and the grammar of attitude reports. *Semantics and Pragmatics* 19(5). https://doi.org/10.3765/sp.19.5.

This version will be replaced with the final typeset version in due course.
Note that page numbers will change, so cite with caution.

EARLY ACCESS

Direction of fit and the grammar of attitude reports*

Thomas Grano
Indiana University

Abstract Crosslinguistically, 'believe' and related verbs often have complement clause syntax different from that of 'want' and related verbs, although the reasons behind this pattern remain poorly understood. Moltmann (2024) suggests an avenue to explore: the possibility that finite complements apply to attitudes that have a word/mind-to-world direction of fit whereas nonfinite complements apply to attitudes that have a world-to-word/mind direction of fit. In this reply, I show that Moltmann's suggestion faces apparent challenges from *hope* and from fiction predicates like *pretend* and *dream*, although it may be possible to overcome these challenges via appropriate refinements to Moltmann's direction of fit diagnostics and/or to the hypothesis linking direction of fit to finiteness. I close by asking *why* the grammar of attitude reports might correlate with direction of fit and whether a reduction to truth-conditional properties of attitude reports is possible.

Keywords: direction of fit, attitude reports, finiteness, mood

1 Introduction

Crosslinguistically, complements to 'believe' (and semantically related verbs) are often syntactically different from complements to 'want' (and semantically related verbs): 'believe' complements typically resemble declarative main clauses (Hacquard & Lidz 2019) whereas 'want' complements often take a different shape, at least sometimes involving a smaller structure (Wurmbrand & Lohninger 2023). In English, for example, this manifests as a finite/nonfinite split: *believe* accepts finite complements, as in (1), whereas *want* requires a nonfinite complement, as in (2).

(1) Pat believes [(that) it will rain tomorrow].

(2) Pat wants [(for) it to rain tomorrow].

Despite the prevalence of such patterns, the reasons behind them remain poorly understood. In the concluding chapter of her recent book on the semantics of attitude reports and modals, Moltmann (2024) suggests — as an avenue for further

* I am grateful to *S&P* handling editor Judith Tonhauser and two anonymous *S&P* reviewers for their constructive feedback and encouragement on earlier versions of this article. Of course, any remaining errors are my own.

development — the possibility that finite complements apply to attitudes that have a word/mind-to-world direction of fit whereas nonfinite complements apply to attitudes that have a world-to-word/mind direction of fit. I will henceforth refer to this suggestion as the *direction of fit hypothesis*. Direction of fit is an influential concept in speech act theory (Searle 1969) and philosophy of mind (Searle 1983).[1] Roughly, speech acts like assertions are said to have a word-to-world direction of fit and attitudes like belief a mind-to-world direction of fit in the sense that if the assertion or belief does not match the world, there is normative pressure to reject the assertion or revise the belief, respectively. By contrast, speech acts like directives are said to have a world-to-word direction of fit and attitudes like intentions and desires a world-to-mind direction of fit in the sense that mismatches in these cases involve pressure in the opposite direction: for the addressee (or attitude-holder) to change the world to make the content of the directive (or the intention or desire) true.

Although prominent in some areas of linguistics and philosophy, direction of fit has received little attention as a way of understanding patterns like (1)–(2). A notable exception is Farkas (1992), who entertains but ultimately rejects direction of fit as a way of understanding mood choice in Romance. Farkas concludes, "Conceptually, the main problem with this approach is that the fundamental distinction — that between the two directions of fit — has to be given a more precise characterization if it is to be made operative" (p. 78). A major question addressed in this work is whether Moltmann's characterization of direction of fit is precise enough to overcome this objection.

If the direction of fit hypothesis were to pan out, that would be an interesting result, because it would potentially mean that the grammar of attitude reports goes beyond the capacity of truth-conditional semantics. That is, (non)finiteness in (1)–(2) would not be contributing to or reflecting anything about what makes these sentences true but instead tracking something about the norms associated with the attitudes that these sentences report on. In that vein, Moltmann points out that direction of fit goes beyond what her semantic framework, object-based truthmaker semantics, can account for, and says that therefore, "truthmaker semantics is to be embedded in a richer semantics of mental and illocutionary objects in which normative and causal notions play a role as well" (p. 87). I will return to this issue below.

The goal of this work is to assess the direction of fit hypothesis. I am mostly interested in two questions. First, does Moltmann characterize direction of fit precisely enough to make testable predictions about which predicates take finite complements and which take nonfinite complements? In other words, does it overcome Farkas's criticism articulated above? Second, assuming that the direction of fit hypothesis (in

1 Prior to Searle, the concept of direction of fit was articulated by Anscombe (1957), although as pointed by Humberstone (1992), the term "direction of fit" is traceable — albeit in a somewhat different sense — to Austin (1953).

some form) is on the right track, *why* should it hold, and can it be reduced to more general principles of grammar? By way of preview, my answer to the first question will be that while Moltmann's definition of direction of fit is precise enough to make testable predictions, the direction of fit hypothesis does not accurately capture the complementation behavior of *hope* or of fiction predicates like *pretend* and *dream*. In the case of *hope*, a modest enrichment to the diagnostics for direction of fit may solve the problem, while fiction predicates may call either for a more serious revision to the definition of direction of fit or for an expanded version of the direction of fit hypothesis. As for the second question, I explore some potential avenues for reducing direction of fit to truth-conditional properties of attitude reports, as a potential precursor to an explanation for the patterns.

The organization of the rest of this article is as follows. In Section 2, I review Moltmann's characterization of direction of fit. In Section 3, I introduce some refinements to the direction of fit hypothesis that are needed to give it a fighting chance of success. In Section 4, I show that the refined version of the hypothesis accounts for some otherwise puzzling complementation patterns across the verbs *believe*, *want*, *hope*, and *intend* — provided that, in departure from Moltmann, we classify *hope* as having both mind-to-world and world-to-mind dimensions and we introduce some enrichments to the diagnostics for direction of fit. In Section 5, I turn to fiction predicates, and discuss their implications for the direction of fit hypothesis. Finally, in Section 6, I discuss potential truth-conditional bases for the direction of fit hypothesis.

2 Defining "direction of fit"

Moltmann (2024) articulates a theory of the semantics of attitude reports and modals centered around the concepts "attitudinal objects" and "modal objects", respectively (collectively: "satisfiable objects"). Attitudinal objects include "claims, judgments, beliefs, assumptions, hopes, requests, decisions, desires, intentions, ideas, and hypotheses" (p. 1), while modal objects include "obligations, permissions, laws, rules, offers, invitations, abilities, strategies, options, dispositions, and essences" (p. 1).

Moltmann's theory has many noteworthy features that go beyond the scope of this reply. For context, I mention three here. First, propositions play no role in the theory; instead, complement clauses act as predicates of satisfiable objects, and the relevant satisfiers are not possible worlds (as in possible worlds semantics) but rather situations and actions. This is couched in a framework that Moltmann calls "object-based truthmaker semantics", which is a variation on Fine's (2017) (sentence-based) truthmaker semantics. Second, neither modal nor attitude sentences involve quantification. Instead, characterizing the content of a modal or attitude sentence is a matter of specifying its exact verifiers and (if there are any) exact falsifiers. The

difference between possibility and necessity is captured not in terms of existential vs. universal quantification but instead in terms of whether the satisfiable object in question has verifiers only (possibility) or has both verifiers and falsifiers (necessity). Third, all attitude reports and modal sentences are underlyingly complex, involving a noun referring to a satisfiable object connected via a light verb to the subject, so that sentences like (3a–b) have syntactic analyses that are more transparently reflected in their overtly complex counterparts in (4a–b).

(3) a. Ann believes that it is raining.
b. Ann is permitted to leave.

(4) a. Ann has the belief that it is raining.
b. Ann has permission to leave.

Initially, Moltmann (2024) classifies satisfiable objects into four categories depending on the kinds of satisfaction predicates that apply to them.[2] Truth predicates like *true*, *false*, and *correct* apply to objects that have a word/mind-to-world direction of fit. Predicates of fulfillment (*fulfill*, *satisfy*) and violation (*violate, ignore*) apply to objects that have both a world-to-word/mind direction of fit and the force of necessity. Predicates of acceptance (*accept*, *take up*) apply to objects that have both a world-to-word/mind direction of fit and the force of possibility. Finally, predicates of realization (*realize*, *carry out*) apply to objects like intentions and decisions, which belong to a third and final class of objects that have a world-to-word/mind direction of fit. Examples of each category are given in (5).

(5) a. The belief/assertion was true/false/correct.
b. The request/demand was fulfilled/satisfied/violated/ignored.
c. The offer/invitation was accepted/taken up.
d. The intention/decision/promise was realized/carried out.

Moltmann eventually assigns a special status to the predicate *correct* in her characterization of direction of fit. Specifically, she says, "The attitudinal objects to which *correct* when conveying just truth applies, acceptances, are just the attitudinal objects that come with a word/mind-to-world direction of fit, that is, attitudinal objects whose content ought to fit the world, rather than the other way around. Truth as a norm is then associated with all attitudinal objects with a word/mind-to-world direction of fit" (p. 80).

This leads Moltmann to the following characterization of direction of fit:

2 See also Moltmann 2021 for an important precursor to some of the ideas summarized here regarding satisfaction predicates and direction of fit.

(6) Characterization of direction of fit

a. An attitudinal object *d* has a *word/mind-to-world direction of fit* just in case *d* satisfies its intrinsic norm ('is correct') in a world *w* iff *w* makes *d* true.

b. An illocutionary object *d* has a *world-to-word/mind direction of fit* just in case any action *a* performed in recognition of *d* satisfies the norm imposed by *d* ('is correct') in a world *w* iff *a* is part of *w* and satisfies *d*. (Moltmann 2024:83)

By (6a), beliefs (for example) have a mind-to-world direction of fit because a belief is correct if and only if it is true. By (6b), requests (for example) have a world-to-word direction of fit because any action performed in recognition of a request is correct if and only if it satisfies the request. Moltmann restricts the definition in (6b) to "illocutionary objects" because some nonillocutionary attitudes including hopes and desires don't quite fit the definition. But according to Moltmann, hopes and desires nonetheless have a world-to-mind direction of fit because they "imply a positive emotive response to their satisfaction (under normal circumstances), and reaching that positive response requires for a part of the world to make such an attitudinal object true" (p. 85).

Although Moltmann initially classifies *true*, *false*, and *correct* together as satisfaction predicates that apply to attitudes that have a word/mind-to-world direction of fit, it later becomes clear that these predicates do not behave uniformly with all attitudes. Moltmann acknowledges this in footnote 19 on p. 79, where she observes that some attitudinal objects including guesses, hypotheses, assumptions, answers, impressions, and thoughts can be said to be *correct*, but it sounds odd to call them *true*. Moltmann suggests that this may be because *true* "requires warrant in addition to truth". Conversely, Moltmann points out that stories, propaganda, and publicity can be said to be *true* but not *correct*. About these latter cases, Moltmann says, "A plausible reason is that the aim of the story is not truth, but, say, entertainment. This would also hold for things like propaganda and publicity, attitudinal objects, in a sense, whose aim is not truth, but influencing the mental state of the audience" (p. 79, footnote 19). Thus, *correct* is a reliable diagnostic for a word/mind-to-world direction of fit in Moltmann's sense, whereas *true* is not.[3] This is in keeping with the special status that Moltmann assigns to (normative) "correctness" in her characterization of direction of fit.

3 What, an anonymous reviewer asks, is a reliable diagnostic for a world-to-word/mind direction of fit? As a working hypothesis, I take it that if a (nominalized) attitude predicate accepts any of the satisfaction predicates exemplified in (5b)–(5d) above, then it has a world-to-word/mind direction of fit.

Let me emphasize here the promise that this set-up holds for overcoming Farkas's concern mentioned in Section 1 above that direction of fit is not precise enough to be made operative. Consider the attitudes of assuming and supposing. Relying on an informal conceptual definition of the word/mind-to-world direction of fit as characterizing an attitude that should be revised or rejected when its content does not match the world, we might think that assuming and supposing do not have the word/mind-to-world direction of fit. After all, I can assume or suppose something for the sake of argument or even to engage in counterfactual reasoning, in which case it would defeat the purpose of the attitude to reject or revise it in light of facts about the world. And yet there is a sense in which, in assuming or supposing something, I am temporarily acting as though the world is a particular way. Does that suffice to count as having the word/mind-to-world direction of fit? How can this be adjudicated? Moltmann's *correct* diagnostic is decisive: to say that an assumption or a supposition is *correct* is to convey that its content is true, which is the telltale sign of the word/mind-to-world direction of fit in Moltmann's sense. This contrasts sharply with saying, for example, that a request, desire, or intention is correct, which has nothing to do with the truth of its content.[4] And indeed, the predicates *assume* and *suppose* accept finite complements, as in (7), and reject nonfinite complements (at least when those nonfinite complements are introduced by *for* — see Section 3 below), as in (8), which bodes well for the direction of fit hypothesis.

(7) Pat assumes/supposes that it is raining.

(8) *Pat assumes/supposes for it to be raining.

In what follows, we will consider whether the direction of fit hypothesis holds up against a broader range of predicates. First, though, let's get clearer on exactly how to formulate the hypothesis.

3 Articulating and refining the hypothesis

Here is an initial formulation:

4 Strictly speaking, the *correct* diagnostic applies only to attitudes lexicalized as nouns, but I take the following data to suggest both that *correctly* can be used to test direction of fit for attitudes lexicalized as verbs and that direction of fit remains stable across the verb/noun divide.

(i) Pat (correctly) believed/assumed/supposed that the meeting would begin soon.

(ii) Pat (??correctly) requested/wanted/intended for the meeting to begin soon.

(9) *Direction of fit hypothesis (v. 1 of 2)*: If an attitude predicate has a word/mind-to-world direction of fit, then it accepts finite complements, and if an attitude predicate has a world-to-word/mind direction of fit, then it accepts nonfinite complements.

(9) is formulated so that it makes no predictions about attitudes predicates that lack a direction of fit altogether. For example, I assume that factive predicates like *regret* — by virtue of presupposing the truth of their complement — lack a direction of fit (see, e.g., Searle's 1975:356–357 similar take on expressive speech acts). We should therefore understand the direction of fit hypothesis as something to ultimately be embedded in or derived from a more comprehensive theory of complementation. In Section 5 below, I briefly return to this limitation of the hypothesis and consider one potential expansion.

In order for the hypothesis as formulated in (9) to have any chance of success, we first need to refine "finite complements" and "nonfinite complements". "Finite complements" in particular needs to be narrowed to "finite *indicative* complements". The reason is that English also has finite *subjunctive* clauses that typically function as complements to predicates like *request* that have a world-to-word direction of fit:

(10) I request that you be here by noon.

"Nonfinite complements", in turn, needs to be narrowed to "*for-to* complements", in a sense to be made precise. This is because, although many nonfinite-complement-taking predicates have a world-to-word/mind direction of fit, as in (11), there are also some that do not, as in (12).

(11) a. Pat wants to be a doctor.
b. Pat begged to go to medical school.

(12) a. Pat claims to be a doctor.
b. Pat believes her son to be a doctor.

However, the superficial similarity of the nonfinite complements across (11) and (12) obscures an important difference: only those in (11) can be introduced by complementizer *for* when they have an overt subject:

(13) a. Pat wants (very much) for her son to be a doctor.
b. Pat begged for her son to go to medical school.

(14) a. *Pat claims for her son to be a doctor.
b. *Pat believes for her son to be a doctor.

As observed by Bresnan (1972), Pesetsky (1992), Portner (1997), Grano (2016), and others, *for-to* infinitives (with or without overt *for*) are distinct from other

kinds of infinitives, and in the tradition of these previous researchers, I define a *for-to* complement as a nonfinite complement that can (and in many cases must) be introduced by complementizer *for* when its subject is overt.

Given these considerations, I revise the hypothesis as follows:

(15) *Direction of fit hypothesis (v. 2 of 2)*: If an attitude predicate has a word/mind-to-world direction of fit, then it accepts *finite indicative* complements, and if an attitude predicate has a world-to-word/mind direction of fit, then it accepts *for-to complements*.

This hypothesis is couched in English-specific terms (especially with its reference to *for-to* complements), but as alluded to in the introduction, it is crosslinguistically common for languages to syntactically mark a difference between complements to 'believe'-like predicates vs. 'want'-like predicates, although the details of this marking vary from one language to the next (cf. also Bolinger 1968). And indeed, the suggestion from Moltmann (2024) on which this hypothesis is based is offered in an explicitly crosslinguistic spirit in that she prefaces it with the observation — credited to Wurmbrand & Lohninger (2023) — that "some languages . . . display a a general semantic distinction between the choice of finite clauses and infinitival clauses as complements" (p. 209). In what follows, I will mostly confine myself to English, in order to keep the evaluation of the hypothesis manageable within the confines of this reply. But I am optimistic that the considerations brought forth below may have analogues in other languages, given what we already know about crosslinguistic similarities in complementation patterns.

4 Dataset #1: Believing, wanting, intending, and hoping

The pattern in (16)–(17) will provide our initial testing grounds for the direction of fit hypothesis:

(16) Pat believes/*wants/*intends/hopes that her son is a doctor.

(17) Pat *believes/wants/intends/hopes for her son to be a doctor.

In (16)–(17), we observe that *believe* accepts finite indicative complements only, *want* and *intend* accept *for-to* complements only, and *hope* accepts both complement types.

I begin by briefly reviewing a recent account of this pattern. My goal in doing this is to illustrate that the pattern presents a genuine challenge, and it would be a credit to the direction of fit hypothesis if it were able to account for it in terms that are simpler than currently available alternatives. Specifically, this pattern was recently analyzed by Grano (2024), who builds on Portner & Rubinstein 2020 and

others regarding mood choice with 'hope' in Romance. Across Romance, while 'believe' is a stable indicative-selector (with some exceptions) and 'want' a stable subjunctive-selector, 'hope' is more variable. For example, it is a subjunctive-selector in Spanish but typically an indicative-selector in French. To address this, Portner & Rubinstein (2020) propose that indicative clauses function as complements to attitude predicates that involve one modal background whereas subjunctive clauses function as complements to attitude predicates that involve two modal backgrounds. 'Believe' accepts indicative complements because it involves one (doxastic) modal background; 'want' accepts subjunctive complements because (following Heim 1992 and others) it involves two modal backgrounds, one doxastic(-like) and the other bouletic. As for 'hope', their proposal is that it is like 'want' in involving two modal backgrounds, but unlike 'want', it imposes requirements on those backgrounds that allow them to be unified into one on a language-specific basis, giving rise to variable mood choice. Those requirements are: realism with respect to the subject's beliefs ((18a), cf. (18b)) and internal consistency ((19a), cf.(19b)).

(18) a. ??Pat hopes to live forever.
b. Pat wants to live forever.

(19) a. ??Pat hopes to leave and hopes to stay. (He can't make up his mind!)
b. Pat wants to leave and wants to stay. (He can't make up his mind!)

Grano (2024) shows that this account makes the wrong predictions about 'intend': 'intend' patterns like 'want' in rejecting indicative complements in English and across Romance, but it patterns like 'hope' in terms of realism and internal consistency:

(20) ??Pat intends to live forever.

(21) ??Pat intends to leave and intends to stay. (He can't make up his mind!)

Grano argues that 'intend' has a property distinguishing it from both 'want' and 'hope' that may help explain its syntax: its content involves causal self-reference, in that if an agent intends to do something, the content of that intention is that that intention lead, in the right way, to that outcome (Harman 1976, Searle 1983). Grano proposes that — because causation is a relation between eventualities — implementing causal self-reference compositionally requires that the complement clause instantiate eventuality abstraction, that is, a lambda-bound eventuality argument that can be manipulated by a CAUSE predicate incorporated into the semantics of *intend*. Crucially, Grano furthermore proposes that eventuality abstraction is available to subjunctive and *for-to* complements but not to finite indicative complements. Toward a unified theory of mood choice, Grano proposes that Romance subjunctive complements, as well as English *for-to* complements, function to signal a "departure from a default

setting" (*à la* von Fintel & Iatridou 2023:1505), the default in this case being the clausal semantics associated with unembedded assertions. One such departure is the presence of two modal backgrounds instead of just one (*à la* Portner & Rubinstein 2020), and another is the presence of eventuality abstraction, as in the case of 'intend'. Following Portner & Rubinstein (2020), the two modal backgrounds associated with 'hope' can sometimes be unified into one, leading to variation, but in the case of 'intend', such unification does not enable indicative complements, because of the need for eventuality abstraction.

Does the direction of fit hypothesis offer another way of understanding the pattern? According to Moltmann, beliefs have a mind-to-world direction of fit whereas wants (desires), hopes, and intentions all have a world-to-mind direction of fit. This is borne out by data like (22): only in the case of belief does *correct* indicate nothing more and nothing less than the truth of the content of the attitude.

(22) Ann's belief/??desire/??hope/??intention was correct.

By the *correct* diagnostic, all the data in (16)–(17) are predicted by the direction of fit hypothesis except that *hope* unexpectedly allows finite indicative complements. However, I would like to propose another direction of fit diagnostic that gives more nuanced results. Word/mind-to-world attitudes, by definition, are supposed to fit the world. Therefore, a good reason for holding a word/mind-to-world attitude would be because that attitude does — or is likely to — fit the world. Accordingly, the sentence variants in (23) are interpreted to mean that Ann's belief, assertion, or assumption, respectively, is likely to be satisfied.

(23) Ann has good reason to believe/assert/assume that she will be a doctor one day.

By contrast, good reasons for holding world-to-word/mind attitudes do not have to do with their likelihood of being satisfied (or at least, that is not their *primary* rationale), and so sentence variants like (24) — to the extent that they are felicitous at all — are not interpreted to mean that Ann's desire, intention, or promise, respectively, is likely to be satisfied.[5]

(24) ??Ann has good reason to want/intend/promise to be a doctor one day.

5 Some variants of (24) may have felicitous uses — consider, for example, (i), supplied by an anonymous reviewer.

(i) Coming from poverty, Ann has good reason to want to become rich.

Crucially, however, (i) cannot be interpreted to mean that Ann's desire is likely to be satisfied.

Let's call this the *good reason* test: for any attitude predicate A, if the frame "[subject] has good reason to A [complement clause]" can convey that the content of the attitude is or is likely to be satisfied, then A has the word/mind-to-world direction of fit. Now for the punchline: when *hope* takes a finite indicative complement, as in (25a), it patterns like the attitudes in (23),[6] and when it takes a *for-to* complement, as in (25b), it patterns like the attitudes in (24).

(25) a. Ann has good reason to hope that she will be a doctor one day.
b. ??Ann has good reason to hope to be a doctor one day.

This suggests not only that hopes have both mind-to-world and world-to-mind dimensions, but also that the finiteness of the complement to *hope* plays some role in how these dimensions are highlighted or activated.[7]

We also see this effect in anankastic conditionals. As observed by Grano (2019, 2024), when *hope* appears in the antecedent of such a conditional, only nonfinite complementation is felicitous:

(26) a. If you hope to become a doctor one day, you have to get good grades.
b. ??If you hope that you will become a doctor one day, you have to get good grades.

This is because anankastic conditionals express a necessary condition for achieving some goal specified in the antecedent (Sæbø 2001), and goals have a world-to-mind direction of fit.

Complicating the picture is the observation that the finiteness of the complement has no effect on what kinds of satisfaction predicates are felicitous with *hope* — whether the complement is finite or not, only world-to-mind satisfaction predicates work:

(27) Ann's hope to become a doctor was fulfilled/??true/??false/??correct.

(28) Ann's hope that she would become a doctor was fulfilled/??true/??false/??correct.

6 I owe this observation and its significance in part to Schueler (1991), who says that, "Classifying hope as having world-to-mind direction of fit covers up or ignores . . . belief-like features of hope, such as the fact that such phrases 'grounds for hope that *p*' or 'good reason to hope that *p*' typically refer to exactly the same things as would such phrases as 'grounds for belief that *p*' or 'good reason to believe that *p*'" (p. 279).

7 An anonymous reviewer reports the intuition that (25a) has two readings, one according to which "Ann's good reason for hoping is about . . . desirability of being a doctor", as brought out by the felicity of a continuation like . . . *even though she thinks it is unlikely* (cf. (i) in note 5 above), and the other according to which Ann's good reason for hoping is about the likelihood of her hope being realized. This complicates the relationship between finiteness and direction of fit, but crucially, the two nonetheless interact in that the latter reading is unavailable to (25b).

That said, *false* does work with *hope* as an attributive adjective, as in *to give someone false hope* or *to have false hope*. To have false hope is to hope for some outcome despite its unlikelihood or impossibility. Here, *false* deals with the mind-to-world dimension of hope. Contrast this with *false intention*: to have a false intention is not to intend for some outcome despite its unlikelihood or impossibility. Rather, a false intention is not an intention at all. To have a false intention is to pretend to have an intention. This contrast in the behavior of *false* between *hope* and *intend* provides further evidence for the view that although *hope* and *intend* may both have a realism requirement (see (18a) and (20) above), only *hope* has a genuine mind-to-world dimension.

While it is beyond the scope of this reply to work out a formal analysis of *hope*'s hybrid direction of fit and how that hybrid status interacts with finiteness, these considerations seem to bode well for the direction of fit hypothesis as a viable way of understanding the puzzling complementation pattern across *believe*, *want*, *hope*, and *intend*. This result does come at a cost, however: given the behavior of *hope*, we can no longer maintain the *correct* test as a fool-proof diagnostic of the word/mind-to-world direction of fit. Does this threaten the project of rendering direction of fit precise enough "to be made operative" (to echo Farkas as quoted in Section 1 above)? Not necessarily, because the facts that we have seen so far are consistent with two possibilities. One possibility is that the *correct* test does work as a fool-proof diagnostic for predicates that have a single direction of fit only, and to test for whether a predicate has a hybrid direction of fit, we need to resort to the *good reason* test. Another possibility — not mutually exclusive with the first — is that the *good reason* test alone does the job of diagnosing the word/mind-to-world direction of fit whether the predicate in question has both directions of fit or not. In either or both of these scenarios, we are still in a position — when coupled with the direction of fit hypothesis — to make testable predictions about the syntax of an attitude predicate's complement.

Let me bring *fear* into the picture as well. Anand & Hacquard (2013) collectively refer to *hope* and *fear* as "emotive doxastic" predicates because they both have a hybrid semantics involving both a preference component and a belief component. They are mirror images of each other in terms of the preference component in that *hope* conveys preference while *fear* conveys dispreference, but when it comes to the belief component, both work in such a way that *x hopes/fears p* conveys that *x* believes that *p* is a possibility. This suggests that *fear* should pattern like *hope* in having a hybrid direction of fit, which in turn leads to the prediction — via the direction of fit hypothesis — that it should accept both finite indicative and *for-to* complements.

Initially, Moltmann (2024) classifies *fear* as having a world-to-word/mind direction fit (p. 81) (she notes in particular that it does not pass the *correct* test), but (29)

shows that *fear* does pass the *good reason* test for word/mind-to-world direction of fit, which supports the idea that it is indeed like *hope* in having a hybrid direction of fit.

(29) Ann has good reason to fear that she will fail her exams.

As far as complement syntax goes, however, *fear* is like *hope* in accepting finite indicative complements, but unlike *hope* in rejecting *for-to* complements, seemingly against the expectations of the direction of fit hypothesis:

(30) Ann fears that her son is in danger.

(31) *Ann fears for her son to be in danger.

This asymmetry between *hope* and *fear* may be related to another asymmetry that Moltmann points out, namely that unlike *hope*, *fear* does not accept the satisfaction predicate *be fulfilled*:

(32) ???John's fear that he would lose was fulfilled. (Moltmann 2024:85)

According to Moltmann, this is because "[t]he positive emotive response that a hope is directed toward constitutes a kind of norm or purpose and as such imposes a requirement on the world, rather than being subject to a requirement itself. Negative emotive responses that go along with a fear becoming true do not seem to be able to set up such a norm or purpose" (p. 85). Perhaps, then, this lack of norm or purpose means that *fear* does not have a world-to-word/mind direction fit after all, which would mean that data like (31) are compatible with the direction of fit hypothesis. If *fear* has a word/mind-to-world direction of fit only, it spells trouble for the *correct* test as being diagnostic of such predicates (because *fear* fails this test), but the *good reason* test is still in the running as a fool-proof diagnostic of word/mind-to-world direction of fit.

5 Dataset #2: Fiction predicates

Now let's consider a class of predicates that Farkas (1992) calls "fiction verbs", which include *pretend*, *fantasize*, *dream*, and *imagine* (on one of its usages[8]). These verbs consistently accept finite indicative complements while rejecting *for-to* complements:

(33) I pretended/fantasized/dreamed/imagined that we were in a movie.

(34) *I pretended/fantasized/dreamed/imagined for us to be in a movie.

8 Not relevant here is the sense of *imagine* that is roughly synonymous with *believe*, as in *I imagine it might rain tomorrow*.

Notably, though, these predicates fail both the *correct* test and the *good reason* test for the word/mind-to-world direction of fit; that is, to the extent that they are felicitous at all, the sentences in (35) do not convey that the content of Ann's pretense (or fantasy, dream, etc.) is true, nor do the sentences in (36) convey that the content of Ann's pretense (or fantasy, dream, etc.) is true or likely to be true.

(35) ??Ann's pretense/fantasy/dream/imagination was correct.

(36) ??Ann has good reason to pretend/fantasize/dream/imagine that she is in a movie.

If the direction of fit hypothesis as articulated in (15) above is to be maintained, this means that either (i) fiction predicates lack a direction of fit altogether and are therefore outside the scope of the hypothesis or (ii) fiction predicates do have the word/mind-to-world direction of fit but there is a problem with our diagnostics for direction of fit. Here I will consider each possibility in turn.

If fiction predicates lack a direction of fit, then the data above do not threaten the direction of fit hypothesis; however, the more predicates there are that fall outside the scope of the hypothesis, the weaker that hypothesis is. In Section 3 above, I already set factive predicates aside. If fiction predicates are to be set aside too, the hypothesis is weaker still. Ideally, fiction predicates' demand for finite indicative complements ought to follow from the same general principles as that of word/mind-to-world predicates. In that spirit, let me entertain an expansion of the direction of fit hypothesis. I am grateful to an anonymous reviewer for suggesting the following way of fleshing out this idea. According to an influential view in philosophy, fiction involves pretense (see Kroon & Voltolini 2024 for an overview). Suppose, in this vein, that fiction predicates are semantically complex in the sense that they involve a "pretense" modifier operating on an attitude that does have the word/mind-to-world direction of fit. For example, *pretend that S* would have an analysis paraphrasable as *pretend to accept/believe/assume that S*, *imagine that S* would be something like *pretend to believe through visualization that S* (both of these paraphrases come from the reviewer), and *dream that S* would involve, according to the reviewer, "some form of passive pretense". Then, we might expand the direction of fit hypothesis as follows:

(37) *Expanded direction of fit hypothesis:* If an attitude predicate has a word/mind-to-world direction of fit *or is composed of a pretense modifier operating on an attitude with a word/mind-to-world direction of fit*, then it accepts finite indicative complements, and if an attitude predicate has a world-to-word/mind direction of fit, then it accepts *for-to* complements.

The crucial question is: is this hypothesis still precise enough to make testable predictions? How do we diagnose attitude predicates that are composed of a pretense modifier operating on an attitude with a word/mind-to-world direction of fit? One idea might be: can we felicitously paraphrase the attitude in a way that explicitly uses an attitude predicate with the word/mind-to-world direction of fit? The predicate *pretend* seems ready-made for such a test in that it can indeed be felicitously paraphrased as *pretend to believe*. It is also striking that the overtly complex expression *make believe* is a standard synonym of *pretend*.

This diagnostic does not extend to all fiction predicates, however. For example, a sentence like *I dreamed I was falling* cannot be felicitously paraphrased as *I dreamed that I believed that I was falling* or any other variation that I can think of. An anonymous reviewer points out that there is debate in philosophy over whether dreaming is like imagination or like perception (see, e.g., Windt 2021 for an overview) and that in the latter case, we might expect *dream* to pattern like perception predicates (which are factive) in relevant respects and therefore lack a direction of fit. If that is on the right track, then the expanded direction of fit hypothesis covers some fiction predicates, whereas others are like factives in being beyond its scope.

The second option is to maintain the simpler version of the direction of fit hypothesis and instead revise our understanding of direction of fit in such a way that fiction predicates have the word/mind-to-world direction of fit after all. There is in fact such an approach to direction of fit on the market: Archer (2015) proposes to reconceive direction of fit so that an attitude has a mind-to-world direction of fit just in case it is truth evaluable, as diagnosed by its ability to participate in classically valid inferences. Archer gives the example of pretending to be Harry Potter while attending a fantasy convention. In that context, the premises (38a) and (38b) entail the conclusion (38c) even though (38a) is a mere pretense and I believe neither (38a) nor (38c).[9]

(38) a. I am Harry Potter.
b. If I am Harry Potter, then Voldemort killed my parents.
c. Therefore, Voldemort killed my parents. (Archer 2015:176)

Attitudes with a world-to-word/mind direction of fit, by contrast, cannot participate in classically valid inferences, as brought out most dramatically by directive speech

9 In a similar vein, Farkas (1992) — although she ultimately rejects a direction of fit approach to complement syntax; see Section 1 above — considers fiction predicates to have a mind-to-world direction of fit because they "describe a world" (p. 77) "even though the world they describe happens not to be the actual world" (p. 77).

acts. For example, (39a) and (39b) cannot be said to entail (39c) because (39a) is not truth evaluable.[10]

(39) a. Shut the office door!
b. If the office door is shut, then professor Smith is away.
c. Therefore, professor Smith is away. (Archer 2015:174)

Again here, though, the crucial question is: is this characterization of direction of fit precise enough to make testable predictions? As pointed out to me by an anonymous reviewer, (38) may involve a situation of pretending, but we cannot conclude anything from it about the predicate *pretend*, given that that predicate does not appear anywhere in the example.

In summary, the complementation behavior of fiction predicates is apparently not accurately predicted by the direction of fit hypothesis. We might rectify this situation by expanding the hypothesis or by revising the notion of direction of fit, but both options raise new questions about how to run tests on predicates in a way that allows us to predict the syntax of their complements.

6 *Why* should direction of fit correlate with finiteness?

Assuming that the direction of fit hypothesis (in some form) is correct, *why* should it hold? That is, how can we integrate the hypothesis with — or derive it from — general principles of grammar?

I assume as background, in line with mainstream generative theories, that natural language grammar contains both a syntactic module that generates sentences and a semantic module that interprets them (see, e.g., Heim & Kratzer 1998). Strings may be unacceptable either because they are not generated or because they give rise to an interpretive problem. (See Abrusán 2019 for an overview.) I furthermore take the position that when an acceptability pattern has a semantic distribution (such as is famously the case for the distribution of negative polarity items or the distribution of strong vs. weak quantifiers in *there*-sentences, or — the case at hand — the relative distribution of finite indicative vs. *for-to* clauses), this lends strong initial credence to the hypothesis that the pattern has to do with the interpretive component of the grammar.

In the introduction, I said that if the grammar of attitude reports is indeed sensitive to direction of fit, this would be an interesting result, because it would potentially

10 This is not to say that directives have no logic of their own — Archer (2015) entertains the idea that what he calls "practical attitudes" may be able to participate in certain kinds of inferences. Rather, (39) is intended to show that directives cannot participate in arguments that are valid in the "classical sense", whereby "an argument is valid just in case the truth of its premises guarantees the truth of its conclusions" (p. 177).

place that grammar outside the confines of truth-conditional semantics, insofar as direction of fit has to do with the norms associated with different attitudes rather than with the truth conditions of the sentences that report on these attitudes. The keyword here, though, is *potentially*, because it could turn out that direction of fit, even if not a truth-conditional concept itself, correlates with something that is truth-conditional and that is ultimately responsible for the pattern. In what follows, I will run through some possible candidates for such correlates. Of course, finding a truth-conditional correlate of direction of fit would not by itself *explain* the patterns. But it might give us some idea about where to seek explanations. For example, according to one productive line of research, certain kinds of sentences are unacceptable because they have grammatical properties that make them logically trivial (tautologous or contradictory) — see, for example, Chierchia 2013 for an approach to polarity phenomena in this vein. Because triviality is a matter of truth conditions (i.e., truth under *all* or *no* conditions), pursuing such an approach to complementation would require a characterization of the patterns in truth-conditional terms.[11]

In the context of Moltmann's theory, one place to look for a truth-conditional correlate of direction of fit would be the situation/action distinction. Indeed, right before suggesting the connection between direction of fit and finiteness, Moltmann (2024:209) observes, "In certain cases the choice of a finite or an infinitival clause indicates a difference in the satisfiers of the respective attitudinal objects: the finite clausal complement applies to an attitudinal object whose satisfiers are situations, the infinitival complements applies to one whose satisfiers are actions (*John decided that he was sick* as opposed to *John decided to leave*)." Thus we might pursue the hypothesis that finite indicatives denote, on some level, sets of (nonactional) situations, whereas *for-to* clauses denote sets of actions, and are therefore subject to different compatibility requirements with their embedding predicates. Unfortunately, however, as Moltmann (2024) points out in a different context, certain attitudinal objects, including hopes and desires, have a world-to-word/mind direction of fit, "even though they do not require actions to be their satisfiers" (p. 85). (Consider, e.g., *John wants it to be sunny tomorrow*.) While there may be a robust ban on finite indicative clauses as actional complements (see Grano 2024 for extensive discussion), not all *for-to* clauses are actional, so a reduction of the direction of fit hypothesis to a situation/action distinction is not ultimately viable.

11 To be sure, Moltmann's theory of attitudes itself does not necessarily give us reason to expect that a reduction to truth-conditional concepts should be possible, insofar as Moltmann treats truth as one sort of norm among a broader family of norms that different content bearers associate with. Nonetheless, I consider this a worthwhile exercise, given the promise that truth-conditional approaches have shown for illuminating other semantically driven acceptability patterns, as alluded to above. I thank an anonymous reviewer for prompting me to clarify this point.

Looking to the literature on mood choice, another possibility — albeit one not necessarily compatible with Moltmann's background theory — would be to try to reduce direction of fit to the distinction between one vs. two modal backgrounds. As discussed in Section 4 above, Portner & Rubinstein (2020) pursue a theory in which Romance indicative clauses function as complements to attitude predicates that involve one modal background whereas subjunctive clauses (which are distributionally similar to English *for-to* clauses) function as complements to attitude predicates that involve two modal backgrounds. There is a conceptual connection between having the world-to-word/mind direction of fit and having two modal backgrounds in the sense that one modal background provides an information state and the other modal background expresses a ranking over that state. That ranking typically encodes a *preference* for some outcomes over others, formally cashing out the concept of a world-to-word/mind direction of fit. But this reduction is also not without its problems. As already discussed in Section 4, it handles *hope* only at the expense of mishandling *intend*. Another potential issue concerns complementation with *possible*. When *possible* takes a *for-to* complement, the result is compatible with an ability ascription interpretation. (E.g., *It is possible for John to lift 300 pounds*.) While abilities arguably have the world-to-mind direction of fit (for example, we speak of abilities being *realized*, and *realize* is a satisfaction predicate that goes with this direction of fit), it is not clear that ability ascriptions involve any sort of ranking.

Yet a third possibility would be to try to reduce direction of fit to differences in temporal properties. In particular, attitude predicates that have a world-to-word/mind direction of fit typically have an obligatory future orientation, for obvious reasons: the past cannot be changed; therefore, any attitude seeking to change the world must look to the future. See especially Portner 1997 for an approach to the distribution of *for-to* clauses along these lines, and see also relevant discussion around p. 85 of Moltmann 2024. One potential selling point of a reduction to temporal properties is that at least some of the few world-to-word/mind predicates that go against the general trend in allowing past orientation are precisely ones that deviate from the expected complementation patterns. In Section 4 above, I explored the idea that *hope* and *fear* accept finite complements because they have a mind-to-world dimension, but it may also be relevant that they allow past orientation; for example, if I don't know whether it snowed overnight, I can *hope* or *fear* that it did. This temporal approach would be threatened, however, if we could find a word/mind-to-world predicate that has obligatory future orientation yet accepts finite indicative complements only. A potential case of this is *foresee*. To the extent that there is a difference in acceptability between *On Christmas Eve, Ann (correctly) foresaw that it would snow on Christmas* and *?On Christmas, Ann (correctly) foresaw that it had snowed on Christmas Eve*, it would seem that direction of fit pulls apart from temporal orientation and that it is direction of fit that drives complementation.

Of course, it is possible that the search for a truth-conditional correlate of direction of fit is in vain, and that we do need semantic mechanisms that go beyond truth conditions in the appropriate ways in order to account for complementation patterns in the grammar of attitude reports. In the absence of clear answers, I will end on this uncertain note, with the hope that this discussion has at least clarified some of the issues that are at stake in this inquiry.

References

Abrusán, Márta. 2019. Semantic anomaly, pragmatic infelicity, and ungrammaticality. *Annual Review of Linguistics* 5. 329–351. https://doi.org/10.1146/annurev-linguistics-011718-011938.

Anand, Pranav & Valentine Hacquard. 2013. Epistemics and attitudes. *Semantics & Pragmatics* 6. 1–59. https://doi.org/10.3765/sp.6.8.

Anscombe, G.E.M. 1957. *Intention*. Oxford: Blackwell.

Archer, Avery. 2015. Reconceiving direction of fit. *Thought* 4(3). 267–316. https://doi.org/10.1002/tht3.171.

Austin, J.L. 1953. How to talk: Some simple ways. In *Aristotelian society*, vol. 53 1, 227–246. https://doi.org/10.1093/aristotelian/53.1.227.

Bolinger, Dwight. 1968. Post-posed main phrases: An English rule for the Romance subjunctive. *Canadian Journal of Linguistics* 14(1). 3–30. https://doi.org/10.1017/S0008413100019629.

Bresnan, Joan. 1972. Theory of complementation in English syntax. Ph.D. Dissertation, Massachusetts Institute of Technology. http://hdl.handle.net/1721.1/12979.

Chierchia, Gennaro. 2013. *Logic in grammar: Polarity, free choice, and intervention*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199697977.001.0001.

Farkas, Donka. 1992. On the semantics of subjunctive complements. In Paul Hirschbühler & E.F.K. Koerner (eds.), *Romance languages and modern linguistic theory*, 69–104. Amsterdam & Philadelphia: Benjamins. https://doi.org/10.1075/cilt.91.07far.

Fine, Kit. 2017. Truthmaker semantics. In Bob Hale, Crispin Wright & Alexander Miller (eds.), *Companion to the philosophy of language*, Chichester, UK: John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118972090.ch22.

von Fintel, Kai & Sabine Iatridou. 2023. Prolegomena to a theory of X-marking. *Linguistics and Philosophy* 46. 1467–1510. https://doi.org/10.1007/s10988-023-09390-5.

Grano, Thomas. 2016. Semantic consequences of syntactic subject licensing: Aspectual predicates and concealed modality. In *Sinn und Bedeutung (SuB)*, vol. 20, 306–322. https://doi.org/10.18148/sub/2016.v20i0.265.

Grano, Thomas. 2019. Belief, intention, and the grammar of persuasion. In *Chicago Linguistic Society (CLS)*, vol. 54, 125–136.

Grano, Thomas. 2024. Intention reports and eventuality abstraction in a theory of mood choice. *Linguistics and Philosophy* 47. 265–315. https://doi.org/10.1007/s10988-023-09397-y.

Hacquard, Valentine & Jeff Lidz. 2019. Children's attitude problems: Bootstrapping verb meaning from syntax and pragmatics. *Mind & Language* 34(1). 73–96. https://doi.org/10.1111/mila.12192.

Harman, Gilbert. 1976. Practical reasoning. *The Review of Metaphysics* 29(3). 431–463. https://www.jstor.org/stable/20126812.

Heim, Irene. 1992. Presupposition projection and the semantics of attitude verbs. *Journal of Semantics* 9(3). 183–221. https://doi.org/10.1093/jos/9.3.183.

Heim, Irene & Angelika Kratzer. 1998. *Semantics in generative grammar*. Malden, MA: Blackwell Publishing.

Humberstone, Lloyd. 1992. Direction of fit. *Mind* 101(401). 59–83. https://doi.org/10.1093/mind/101.401.59.

Kroon, Fred & Alberto Voltolini. 2024. Fiction. In Edward N. Zalta & Uri Nodelman (eds.), *The Stanford encyclopedia of philosophy*, Metaphysics Research Lab, Stanford University Summer 2024 edn. https://plato.stanford.edu/archives/sum2024/entries/fiction/.

Moltmann, Friederike. 2021. Truth predicates, truth bearers, and their variants. *Synthese* 198 (Suppl 2). S689–S716. https://doi.org/10.1007/s11229-018-1814-8.

Moltmann, Friederike. 2024. *Objects and attitudes*. New York: Oxford University Press. https://doi.org/10.1093/oso/9780190878481.001.0001.

Pesetsky, David. 1992. Zero syntax, vol. II. Ms., Massachusetts Institute of Technology.

Portner, Paul. 1997. The semantics of mood, complementation, and conversational force. *Natural Language Semantics* 5. 167–212. https://doi.org/10.1023/A:1008280630142.

Portner, Paul & Aynat Rubinstein. 2020. Desire, belief, and semantic composition: variation in mood selection with desire predicates. *Natural Language Semantics* 28. 343–393. https://doi.org/10.1007/s11050-020-09167-7.

Sæbø, Kjell Johan. 2001. Necessary conditions in a natural language. In Caroline Féry & Wolfgang Sternefeld (eds.), *Audiatur vox sapientiae: A festschrift for Arnim von Stechow*, 427–449. Berlin: Akademie-Verlag. https://doi.org/10.1515/9783050080116.427.

Schueler, G.F. 1991. Pro-attitudes and direction of fit. *Mind* 100(398). 277–281. https://doi.org/10.1093/mind/C.398.277.

Searle, John. 1969. *Speech acts: An essay in the philosophy of language*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139173438.

Searle, John. 1975. A taxonomy of illocutionary acts. In Keith Gunderson (ed.), *Language, mind, and knowledge*, 344–369. Minneapolis: University of Minnesota Press. https://hdl.handle.net/11299/185220.

Searle, John. 1983. *Intentionality: An essay in the philosophy of mind*. New York: Cambridge University Press. https://doi.org/10.1017/CBO9781139173452.

Windt, Jennifer M. 2021. Dreams and dreaming. In Edward N. Zalta (ed.), *The Stanford encyclopedia of philosophy*, Metaphysics Research Lab, Stanford University Summer 2021 edn. https://plato.stanford.edu/archives/spr2026/entries/dreams-dreaming/.

Wurmbrand, Susi & Magdalena Lohninger. 2023. An implicational universal in complementation—theoretical insights and empirical progress. In Horst Lohnstein & Susanne Trissler (eds.), *Propositional arguments in cross-linguistic research: Theoretical and empirical issues*, 183–232. Berlin: Mouton de Gruyter. https://doi.org/10.24053/9783823394105.

Thomas Grano
Ballantine Hall 504
1020 E. Kirkwood Avenue
Bloomington, IN 47405
tgrano@iu.edu