The interaction/satisfaction distinction is redundant:
A reply to Deal (2024) and Oxford (2022)

Christopher Hammerly
Department of Linguistics
University of British Columbia
chris.hammerly@ubc.ca

Preprint Draft – Comments Welcomed – March 15, 2024

Abstract Two recent papers on variation in agreement, one by Deal (2024) for the person-case constraint (PCC) and one by Oxford (2022) for Algonquian direct/inverse Voice, have argued for a split between interaction and satisfaction conditions in Agree (as first proposed in Deal 2015). I show that a simpler probe representation without such a split can capture the same empirical domain covered in these two papers. The model is based in uF features which can be joined on a probe via logical conjunction or disjunction. Matching any subset of a probe's uF features triggers copying, while satisfying the overarching logical statement of the probe via match halts the search. Whether "partial match" between a probe and goal results in deactivation of the subset of matched uF features, or whether deactivation of features only occurs in one fell swoop for all features when the full statement of the probe matches with a goal, is proposed to be a matter of parametric variation. This captures effects otherwise attributed to *dynamic interaction*. When two goals are equidistant to a probe, a Best Match principle decides whether to copy from one or both goals. This renders the distinction between interaction and satisfaction redundant for capturing cross-linguistic variation in patterns of Agree, while preserving other core insights from each account.

1 INTRODUCTION

Agree is a foundational operation in syntax that serves to create redundancies. The potential applications of the operation are vast, but canonically, nominal elements serve as the *goals* for verbal elements (*probes*) looking to copy, share, or cliticize sets of φ-features. Two of the most basic questions that must be answered to have a sufficient theory of Agree are (i) how a probe select goals from which to copy features, and (ii) what directs a probe to halt its search for goals.

One influential answer to these questions has come from the *interaction/satisfaction model*, first proposed in Deal (2015), and recently refined and expanded in Deal (2024). The gist of the model is that probes are specified for two separate sets of conditions. *Interaction* conditions specify which features a probe will copy back from the goal, and thereby which goals the probe will ignore, while *satisfaction* conditions specify which features, once copied from a goal, will halt the probe's search. In the most recent version of the model, the copying of features from certain goals can trigger updates to interaction conditions through a process known as *dynamic interaction*.

The interaction/satisfaction model is squarely posed as a superior alternative to one in which so-called *uninterpretable/unvalued* features (notated *u*F) are the trigger for the application of Agree. In this reply, I argue the opposite: That a model based in *u*F features, which can be understood for present purposes in an abstract sense as features that trigger Agree (e.g. Preminger 2014), has clear conceptual advantages, in that a single set of features can cover the empirical job otherwise done by two separate sets. Therefore the split between interaction and satisfaction (as well as the notion of dynamic interaction) is not necessary within the theory of Agree. I focus on two critical examples, both of which have appeared in the pages of *Linguistic Inquiry* and appear to provide strong support for the interaction/satisfaction model: Deal's 2024 analysis of the person-case constraint (PCC), and Oxford's (2022) analysis of variation in direct/inverse marking in Algonquian. I review the empirical landscape and original approach argued by each of these authors, and show how a *u*F approach cannot be ruled out (and, more strongly, why it should be preferred).

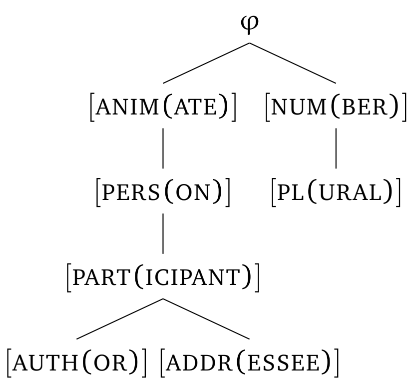
I augment the basic *u*F theory with three key updates, which are already established in the literature or (in some variant) used or acknowledged by Oxford or Deal: (i) *Logically joined probes*: the ability to join features on the probe via logical conjunction and disjunction; (ii) A *deactivation parameter:* a parameter that states whether a probe can deactivate *u*F's on an individual basis when partial match between probe and goal is achieved, or whether features are only deactivated when full match between a probe and goal is achieved; and (iii) *Best Match:* a general principle, which decides which goal(s) to copy features from when two or more goals are equidistant from the probe.

2 INTERACTION AND SATISFACTION

The interaction/satisfaction model originates from Deal (2015), with further refinements presented in Deal (2024, 2022) and Royer and Deal (2023). What distinguishes this model from the many other variants of Agree is the postulation that probes must be defined by two distinct sets of conditions, which can be notated [INT:α, SAT:β], where α and β are sets of features. Interaction conditions (INT) specify which features a probe will copy from a goal. As a result, interaction conditions also dictate which elements a probe will disregard, as any goal that does not match interaction features will not have its features copied back to the probe. The second set, satisfaction conditions (SAT), determine when a probe will stop its search.

As in all current theories of Agree, the feature sets in each condition can vary across individual probes, within limits, accounting for different types of agreement patterns. That is, probes can be *relativized* to certain features (Rizzi, 1990; Béjar, 2003; Preminger, 2014). This has been, for the most part, inextricably tied to the idea that morphosyntactic features are privative features, and particular combinations of features are extrinsically restricted by a feature geometry (Harley and Ritter, 2002; McGinnis, 2005; Béjar, 2003, among many others). For the purposes of convenience and maximum congruency with both Deal (2024) and Oxford (2022), I will assume the feature geometry here, including for the proposed *u*F model.[1] The particular feature geometry relevant to the discussion ahead is given in (1).

(1) *Partial feature geometry*



(2) *Pronominal categories*

a. INAN = {φ}
b. IMPERS = {φ, ANIM}
c. ANIM = {φ, ANIM, PERS}
d. FIRST = {φ, ANIM, PERS, PART, AUTH}
e. SECOND = {φ, ANIM, PERS, PART, ADRR}

This gives rise to the (pro)nominal categories in (2), which define the possible goals for Agree to target that are relevant to the two empirical domains at hand. I restrict the exposition to singular—to make the plural, simply add [Plural] as needed. I also ignore the inclusive/exclusive distinction within the first person and the proximate/obviative distinction within the animate third person, as neither distinction plays a role in either account. The main benefit is that we have a clear way of representing the natural classes that each category falls into. For example, both FIRST and SECOND, the local persons, share [PART] to the exclusion of all other categories, and all categories share the root [φ]. There are many other such relationships that I leave it to the reader to notice.

One key role that the geometry plays in the Deal (2024) version interaction/satisfaction model is to dictate which features come along for the ride when interaction conditions are met. Specifically, Deal (2024) assumes that "*interaction should be understood as copying the designated feature and all features that it geometrically dominates (or: is geometrically entailed by)*" (p. 46). This is essentially what is proposed by Preminger (2014), which he refers to as "snippets" of the feature geometry. It

[1]It is important to note that the feature geometry is flawed in various ways that have been detailed in other work (Nevins, 2007; Harbour, 2016; Cowper and Hall, 2019; Hammerly, 2020, 2021a, 2023) , and will not be repeated here. For the present paper, assuming the feature geometry as, for example, a notational variant of the set-based representation and match algorithm from Hammerly (2021a) is basically harmless, albeit misleading.

must be emphasized (and this is acknowledged by Deal): This does not follow from the logic of the feature geometry itself, where, if F dominates G in the geometry, then the presence of G entails the presence of F, but F does not entail G. Therefore, this idea must be understood as one operating on top of the feature geometry, rather than one that follows from it.

In more recent unpublished work (Royer and Deal, 2023), Deal (with Royer) moves away from this particular conception of interaction, as it is not empirically adequate. The issue is that if the interaction conditions are specified as any of the features along the person branch of the geometry (e.g. [Part] or [Auth]), then number is not expected to be copied, nor any of the geometrically entailed features. Counter to these predictions, there are many cases where the interaction features must be narrowed to a specific person to capture the correct result, but all features seem to be expressed by the morphology on the probe, and therefore all features must be copied to the probe. In order to give the model the best chance possible, I will assume the updated Royer and Deal (2023) version of interaction, where the full set φ-features of a goal that matches the interaction conditions are copied back to the probe. That is, copying is *coarse*. After syntax, the morphophonological particulars of a language will conspire to spell-out forms associated with some subset of these features.

Returning now to satisfaction, it is well-known in the literature on agreement that probes can vary in whether they stop after copying features from the closest goal (e.g. English T agreement, which generally only targets the subject) or whether *overagreement* occurs and features from multiple goals are copied. Cases where the probe uniformly copies features from and stops at the first goal it encounters can be modelled by a probe that interacts with any φ-bearing element and is satisfied when it reaches a φ-bearing element: [INT:φ, SAT:φ]. The more interesting case, and the one that has served as the primary motivation for the model, is overagreement. In particular, patterns of C-agreement in Nez Perce served as the original case study in Deal (2015). Setting aside the morphological particulars, C in Nez Perce copies features from the subject alone, ignoring the object, when the subject is a second person, as exemplified in (3a). However, it copies features from both the subject and object if the subject is anything other than a second person, as exemplified in (3b). In other words, the probe cannot probe past a second person; that is, the probe will only stop once it has reached a second person. In terms of the interaction/satisfaction model, the probe is specified such that it interacts with any φ-bearing element, but is only satisfied by [ADDR]: [INT:φ, SAT:ADDR].

(3) *Nez Perce C-agreement*

a. ke-m kaa *pro* nees-cewcew-teetu *pro*
C-2 then 2SG O.PL-call-HAB.S.SG 1PL
'when you call us' 2SG → 1PL = 2SG

b. ke-pe-m-ex kaa *pro* cewcew-tee'nix *pro*
C-PL-2-1 then 1PL call-HAB.S.PL 2SG
'when we call you' 1PL → 2SG = 1PL + 2SG

The basic distinction between interaction and satisfaction conditions put forward in Deal (2015) remains unchanged in the updated exposition in Deal (2024). The major innovation is a process dubbed *dynamic interaction*, whereby copying certain features from a goal results in an update to the interaction conditions of the probe such that they are more restricted during future probing cycles. Based on other recent work by Deal and collaborators (e.g. Deal, 2022; Royer and Deal, 2023) dynamic interaction is conceptualized as a property of features on goals, not probes. That is, within a given language, certain features on goals are specified for a diacritic that results in the feature being copied specifically to the interaction conditions of the probe, in addition to the probe more generally for the purposes of phonological spell-out. Questions of mechanics and representations aside, the result of dynamic interaction is that, after being open to copying features from any goal, a probe will become pickier about what it copies features from. This process is essential for Deal's account of variation in PCC effects.

3 AN ALTERNATIVE

The proposed alternative is, in many respects, not new: it bears a direct genetic relation to the original model of Agree put forward by Chomsky (2000, 2001). In Chomsky's model, Agree is fundamentally driven by the presence of *unvalued* features (*u*F) on probes. These features must be valued before reaching the PF interface in order to be legible to the morphophonological insertion rules, which rely on values to insert the proper forms.[2] In order to enforce the application of Agree, Chomsky assumes that any leftover *u*F features cause a crash at the interfaces, resulting in ungrammaticality. In subsequent versions of this model, most notably Preminger (2014), this assumption is relaxed: While *u*F features (or their notational equivalent) serve as the trigger for Agree, there are cases where *u*F features are not deleted prior to the interfaces, but the derivation does not crash (a situation dubbed *failed agreement* or *underagreement*). On this revised model, applying Agree is obligatory (and immediate) when a *u*F feature enters the derivation, though its success is not necessary to have a well-formed utterance. This variant of enforcing Agree is the one assumed here: *u*F features obligatorily trigger the application of Agree, and are thus the driver of feature copying, but can simply be ignored if they are not taken care of prior to the interfaces (they do not cause crashes). Ungrammaticality therefore arises in cases where you could or should have agreed, but did not.

We can now turn to detailing how the basic concepts that underlie interaction (triggering feature copying) and satisfaction (halting the probe) can be conceptualized within this framework, where there is only a single set of features governing both processes. Both involve a single principle, *match*, understood within the current representation as feature identity (as in Chomsky, 2000, 2001). If any feature of the probe matches with a feature of the goal, this triggers the copying of the full set of features of the goal to the probe. This is the same assumption made by the updated interaction/satisfaction model presented in Royer and Deal (2023). Satisfaction qua halting a probe's search falls out naturally from the system, with no additional assumptions needed beyond the already necessary idea that *u*F features trigger Agree: Once all *u*F features of a probe have been deactivated, the search for a goal will cease. As well, the probe will halt its search if it has examined all goals

[2] Chomsky additionally assumes that features must be *deleted* prior to LF in order not to be interpreted, as only the features of a goal, not the probe, seem to receive a semantic interpretation. Previous work, such as Hammerly (2019), has argued instead that such features can simply be ignored (e.g. by denoting an identity function).

within its domain, or hits a barrier such as a phase (Chomsky, 2000, 2001) or horizon (Keine, 2020)—this is shared with Deal's system.

One further clarification is needed, and represents, to my knowledge, a novel proposal regarding the parameterization of Agree with a *u*F system (though elements of the proposal have grounding in previous work). I assume that the *partial deactivation* of *u*F features is a matter of parametric variation. Let us use the terms *all-or-nothing* and *some-is-enough* to refer to the two types of deactivation procedures. An all-or-nothing probe will only deactivate *u*F features if the goal provides a *full match* for the features of the probe. A partial match will trigger feature copying, but the matching features will not be disarmed, and therefore stay available to match in future probing cycles. A some-is-enough probe will deactivate any matching features on each probing cycle. So, if there is partial match between the probe and goal, the matching features will *not* be available on further probing cycles. This basically amounts to the idea of *active residue* in Béjar and Rezac (2009), whether the features leftover from an initial cycle of Agree continue to guide further cycles. As we will see in the exposition of the PCC, the two settings of the deactivation parameter will correlate perfectly with cases where dynamic interaction is or is not posited. It can therefore be seen as a sort of replacement principle, which I will argue is conceptually more appealing.

As a proof of concept, let us consider how the system sketched so far fares in capturing the patterns of Nez Perce C-Agreement discussed above in (3). Recall the pattern: The probe agrees with only the subject if the subject is a second person, and otherwise agrees with both the subject and object. So the probe interacts with anything, but is satisfied if it hits a second person. That can be captured here with an all-or-nothing probe [*u*φ, *u*Part, *u*Addr]. This probe will copy all features (critically, including number as well as person), and no features will be deactivated, until a second person is reached; a second person subject will provide a full match, halt the probe, and prevent the copying of features from the object.

Indeed, Deal (2015, 2024) acknowledges that a system without the interaction/satisfaction distinction, such as Béjar (2003), can capture the probing patterns needed for Nez Perce. However, she rejects these competitors on the grounds that they restrict feature copying only to matching features, or those that are geometrically entailed—that is, certain person features. Since Nez Perce C expresses *both* person and number features, these accounts do not capture the particular way that copying occurs. The present account circumvents these issues and criticisms by assuming all feature copying is coarse, and therefore all φ-features come along for the ride when the probe and goal match. As noted numerous times, this is the same assumption made for the interaction/satisfaction model (Royer and Deal, 2023). Indeed, this deflates these original arguments against the *u*F model, since both models can simply make the same assumption regarding copying. Therefore the question becomes whether they can cover the same empirical space. Let us now turn to the two case studies to evaluate this possibility.

4 DEAL 2024: THE PCC

The person-case constraint (PCC) is a person-sensitive phenomenon that restricts certain combinations of clitics in ditransitive constructions, and has been widely studied and debated over the past few decades. As such, a disclaimer is essential: I will adopt all of the same assumptions as Deal does for her analysis of the PCC, except of course for the underlying model of Agree. This is therefore not a wholesale argument against (or even necessarily for) Deal's approach, which has many valuable insights and observations outside of the specific model of Agree that is proposed. I adopt these same assumptions as to have a fair playing field in the assessment of whether or not the interaction/satisfaction distinction is necessary, or whether a model without it based in *u*F features can cover the same ground.

One major point of typological variation in the PCC, and the one that is most pertinent here, is in the exact combinations of direct object (DO) and indirect object (IO) that are ruled in and out. There are five basic classes of PCC effects, as shown in (4) with the use of prominence hierarchies. I assume the reader is familiar with the phenomenon, and Deal provides a clear summary for the uninitiated, so I will not belabour the description here.

(4) *Varieties of the PCC*
The IO must (uniformly) outrank or be equally ranked to the DO on the following scales:

a. *Strictly Descending*: $1 > 2 > 3$
b. *Me-First*: $1 > 2/3$
c. *Weak*: $1/2 > 3$
d. *Strong*: $\{1 > 2 \wedge 2 > 1\} > 3$ (i.e. the DO must be 3)
e. *None:* 1/2/3

These five restrictions are what Deal is able to capture through variation in interaction/satisfaction conditions on the probe, and the specification of dynamically interacting features on goals.

Deal assumes a particular structural relation between the probe and the two goals (IO and DO). Setting aside the speccific cases, which I entirely follow Deal's lead on, the general case of the PCC, and the one we are concerned with here, is one in which the probe first Agrees with the DO, then, if it has not been halted, and the IO matches, it agrees with the IO. I refer the reader to Deal for details on how to achieve this DO preference. In cases where the probe is able to Agree with both goals, both DO and IO clitics are generated, and the relevant utterance is grammatical. Ungrammaticality (or repair) arises when the probe is halted after the initial cycle of Agree with the DO, and thus never probes the IO, so does not generate a clitic.

Let us now look at each particular case. The shape of each subsection is the same: introduce the basic pattern (I refer the reader to Deal and references therein for data and examples), then comparatively detail both Deal's interaction/satisfaction based analysis and the current *u*F analysis. In some cases, there is more than one way for the interaction/satisfaction model to capture a given pattern. I focus just on the specification discussed most by Deal, and refer the reader to the original paper to see the full set of possibilities. The arguments are not affected by this choice.

4.1 *Strong*

The Strong PCC can be described as requiring the direct object to be third person, as summarized in (5). In terms of our model, we need overagreement in the case where a third person is probed first, but and the probe should be stopped after agreement with any local person.

(5) *The Strong PCC*

IO	DO	
1	2	*
2	1	*
1	3	OK
3	1	*
2	3	OK
3	2	*

a. [INT:φ, SAT:PART], no dynamic interaction

b. [*u*φ, *u*Part], Deactivation: all-or-nothing or some-is-enough[3]

Deal's original probe in (5a) will Agree with any DO. If the DO is a third person, it will not be satisfied, so will probe the IO. If the DO is a local person, the search will be halted, and the IO clitic will fail the be generated. The *u*F probe (5b) will also agree with any DO, but will only have leftover features or fail to be deactivated if the DO is third person, since this goal will only match with *u*φ. A local person DO will provide a full match, and therefore fully deactivate the probe and prevent the generation of an IO clitic.

4.2 *Me-First*

The Me-First PCC restricts clitic combinations such that configurations where the DO is a first person are ungrammatical, while all others are allowed. This is summarized in (6). We therefore need the probe to halt if it hits a first person, otherwise it must agree with both the DO and IO.

[3]This parameter setting determines whether the PCC effects generated by probe are simply Strong (3-on-3 allowed) or Superstrong (3-on-3 not allowed; see Haspelmath (2004). Deal adopts an analysis of *3-on-3 cases that is distinct from the PCC via morphological dissimilation, a common approach in the literature. I do not consider these cases further in this reply, as they are not central to the main argument.

(6) *The Me-First PCC*

IO	DO	
1	2	OK
2	1	*
1	3	OK
3	1	*
2	3	OK
3	2	OK

a. [INT:φ, SAT:AUTH], no dynamic interaction

b. [*u*φ, *u*Part, *u*Auth], Deactivation: all-or-nothing

Deal's probe in (6a) will be satisfied when it reaches the first person DO, and therefore fail to generate the IO clitic in those cases, resulting in ungrammaticality. In all other cases it will agree with both the DO and IO. Similarly, the *u*F probe in (6b) will be deactivated only by a first person. Therefore if the DO is a first person, then the probe will be halted before it can generate the IO clitic. All other DOs will fail to fully match the probe, and therefore all features will remain active, allowing all other combinations to be generated.

4.3 Weak

The Weak PCC is a subset of the Strong PCC: only cases where the DO is a local person and the IO a third person are ruled out, while the local-only configurations are grammatical. This is summarized in (7). We therefore need the probe to fail to generate a clitic just in case the IO is a third person.

(7) *The Weak PCC*

IO	DO	
1	2	OK
2	1	OK
1	3	OK
3	1	*
2	3	OK
3	2	*

a. [INT:φ, SAT:–], [Part] dynamically interacts

b. [*u*φ, *u*Part, *u*Auth, *u*Addr], Deactivation: some-is-enough

In Deal's model, this is captured by a so-called *insatiable* probe, as in (7a). These are probes that lack satisfaction conditions, and therefore copy features from all goals in its domain that match its interaction conditions. All else equal, as will be discussed below, such a probe will generate a lack of PCC effects. To get the right behaviour, Deal appeals to dynamic interaction: Initial interaction with a local DO results in the interaction conditions of the probe to be updated to INT:PART. Interaction with a third person DO leaves conditions unchanged. This gets the right result: after agreeing with a first or second person DO, the probe cannot generate a clitic for a third person IO since its interaction conditions have been restricted to [Part]. In all other cases, the probe will successfully generate the local person IO clitic.

The *u*F probe in (7b) is our first some-is-enough probe. Recall that the features of the probe will be deactivated even if there is a partial match with a goal. Going case-by-case, if the DO is a local person, then *u*φ, *u*Part, and *either* *u*Auth or *u*Addr will be deactivated. This leaves only *u*Auth or *u*Addr active for the second cycle with the IO. Therefore, if the IO is a third person, no clitic can be generated and ungrammaticality arises. In the other cases, the IO will always match whatever remains: If the DO was first person, then *u*Addr is left and can match with the second person IO; the converse is true for the second person DO and first person IO. Finally, if the DO is a third person, only *u*φ is deactivated, and the local person IO can generate a clitic through match with *u*Part (and either *u*Auth or *u*Addr). In essence, insatiability can be modelled by having an *overload* of features driving Agree, rather than a *lack* of features. I discuss this further towards the end of the paper.

4.4 Strictly Descending

The Strictly Descending PCC rules out cases where the DO is higher ranked than the IO on the hierarchy 1 > 2 > 3, as summarized in (8). We must therefore halt the probe just in case the DO is a first person with any IO, or the DO is a second person with a third person IO (but allow the IO to be a first person in this case).

(8) *The Strictly Descending PCC*

IO	DO	
1	2	OK
2	1	*
1	3	OK
3	1	*
2	3	OK
3	2	*

a. [INT:φ, SAT:AUTH], [Part] dynamically interacts

b. [*u*φ, *u*Part, *u*Auth], Deactivation: some-is-enough

The interaction/satisfaction probe in (8a) is satisfied if the DO is a first person, therefore cannot interact with the IO to generate a clitic in these cases. If the DO is a second person, dynamic interaction leads the interaction conditions to be updated to INT:PART, therefore it is not possible to generate a clitic if the IO is a third person (but second person is fine). If the DO is a third person, the probe remains the same on the second cycle and can generate a clitic for either a first or second person IO.

The *u*F probe is also be halted by a first person DO, so no clitic from the IO can be generated in these cases. If the DO is a second person, this will partially deactivate the probe, leaving only *u*Auth left. Therefore no third person IO clitic can be generated with a second person DO, but a first person

IO is fine. If the DO is a third person, then both *u*Part and *u*Auth remain active for the second cycle. This allows for the generation of a clitic for either the first or second person IO.

4.5 No effect

When there is no effect, this means that clitics can be grammatically generated in all cases. Therefore, we need a probe that will not be stopped by any possible DO, and will be able to Agree with any IO after the first cycle is complete. We can achieve this on either theory, respectively, with the probes in (9).

(9) a. [INT:φ, SAT:–], no dynamic interaction
b. [*u*φ, *u*Part, *u*Auth, *u*Addr], Deactivation: all-or-nothing

For the interaction/satisfaction theory, an insatiable probe with no dynamically interacting features can do the job. This will interact with and generate clitics for both the DO and IO no matter what the feature combination. On the *u*F theory, a fully-stacked all-or-nothing probe can do the trick. No matter what the features of the DO, it will always match, but it will never match the full set to deactivate the probe. The third person misses *u*Part, *u*Auth, and *u*Addr, the second person misses *u*Auth, and the first person misses *u*Addr. Therefore all features remain active on the second cycle, allowing any IO clitic to be generated as well.

4.6 Summary and a few notes on typology

I showed that a *u*F-based theory of Agree can capture the full range of effects in the PCC with the same assumptions (outside of the model of Agree) that Deal made for the proposed interaction/satisfaction account. In short, the combination of interaction and satisfaction conditions were replaced by a single set of *u*F features (simplifying the representation of the probe) and dynamic interaction was traded for a parameter governing feature deactivation. The relation between probes and PCC effects in the *u*F model is summarized in (10).

(10)	a.	[*u*φ, *u*Part], Deactivation: all-or-nothing or some-is-enough	Strong
	b.	[*u*φ, *u*Part, *u*Auth], Deactivation: some-is-enough	Strictly Descending
	c.	[*u*φ, *u*Part, *u*Auth], Deactivation: all-or-nothing	Me-First
	d.	[*u*φ, *u*Part, *u*Auth, *u*Addr], Deactivation: some-is-enough	Weak
	e.	[*u*φ, *u*Part, *u*Auth, *u*Addr], Deactivation: all-or-nothing	No Effect

Therefore, the proposed system does not undergenerate the PCC typology. But what about other logically possible probes within the system? There are three more possibilities. The first two, given in (11), are those where the probe is specified for *u*φ, *u*Part, *u*Add. These probes predict what Deal refers to as an A-Descending pattern (like strictly descending, by with a hierarchy 2 > 1 > 3) or a You-First pattern (like Me-First, but with second person DOs being ruled out). Deal's system also generates probes that predict these patterns.

(11) a. [*u*φ, *u*Part, *u*Addr], some-is-enough A-Descending
b. [*u*φ, *u*Part, *u*Addr], all-or-nothing You-First

Both of these types of restrictions are, to current knowledge, unattested in the PCC literature. However, both types of agreement patterns are attested outside of the PCC, as discussed in detail by Deal. Therefore, like Deal, I take the absence of these effects in the PCC as an accidental gap in the typology. Regardless, the upshot for the current reply is that both the interaction/satisfaction and *u*F model make identical predictions, so on this front they are on equal footing.

The final logically possible probe is in (12): one with only *u*φ. This probe (whether all-or-nothing or some-is-enough) will agree with and be halted by any φ-bearing element. It is analogous to a [INT:φ, SAT:φ] probe.

(12) [*u*φ], all-or-nothing or some-is-enough

Broadly speaking, this probe is attested: For example, as already discussed, with English T, which agrees with and is halted by the first φ-bearing element it encounters (usually the subject). It can also capture the "indirective" cases discussed by Deal, where the IO is always realized as a non-weak form. We will also see it play a role in the coming discussion of Algonquian.

5 OXFORD 2022: ALGONQUIAN DIRECT-INVERSE

Our second case study comes from Oxford (2022), who provides a comprehensive description and analysis of certain direct/inverse Voice alternations in Algonquian couched in the interaction/satisfaction model. In this study, Oxford focuses on cases of "mixed" alignment where different types of third persons are acting on the local persons (i.e. 3 → 1/2).[4] The languages of the Algonquian family show a striking degree of variation in the inverse in these cases—striking both because languages differ to a significant degree in what particular combinations show inverse, but also because the variation is clearly not random. It follows satisfying clines that beg for a principled explanation. We will see these patterns shortly.

To proceed, we first need a short primer on Algonquian morphosyntax. I will not give a full treatment here, and just focus on the core aspects relevant to the reply. The two probes of interest are those on Voice and Infl, which in Algonquianist terms correspond, respectively, to the *theme sign* and *central agreement*. Consider the two examples from Plains Cree (Central Algonquian) below. Voice shows the aforementioned alternation in direct and inverse forms, with direct always marking the person features of the object (as in the third person theme sign *-a:* below), while inverse is analyzed as a featurally non-specific elsewhere form (*-iko* in the examples below). Infl (*-ya:hk* in the examples below) shows more variation. The key generalization is that Infl always indexes the object when Voice takes an inverse form, and shows either pure subject or subject/object portmanteau

[4] I refer the reader to Oxford for details, but all other cases are "invariant" in some way: non-local only alignments strictly follow a topicality hierarchy, local-only alignments are (almost) never inverse, and mixed alignments with local acting on third is also never inverse.

agreement when Voice is direct. Note, this is the case in the examples below. While the form of Infl is invariant in both, the subject is indexed in (13a) where Voice is direct, while the object is indexed in (13b) where inverse voice arises.

(13) a. wa:pam -a: -ya:hk
see -3OBJ -1PL
'We (EXCL) see her'

b. wa:pam -iko -ya:hk
see -INV -1PL
'She sees us (EXCL)' (Oxford 2022)

Oxford's analysis of these patterns, which will be entirely followed here, is that Voice undergoes impoverishment, leading to the spell out of the elsewhere form qua inverse marker, just in case both Voice and Infl have copied the same set of features via Agree. This explains the correlation between Infl indexing the object and Voice being expressed as the inverse marker: Voice always agrees with *only* the object, so impoverishment is triggered when Infl agrees with *only* the object too. We now have a clear picture of the project at hand: Understand how variation in the probe on Infl leads to different patterns of object agreement across the Algonquian languages, and therefore different patterns of inverse marking.

5.1 Formal tools and asumptions

The first main assumption of Oxford's analysis, which again I follow exactly here, is that the initial agreement relation between Voice and the object leads the features of the object to be available to the probe on Infl. Specifically, since the subject (or agent—I will use the terms interchangeably) is in the specifier of VoiceP, the features of the object on Voice and the subject are *equidistant* to Infl. Given this situation, locality considerations cannot be decisive in the selection of a goal—there are two equally close potential goals to the probe. However, it is clear that Infl *does* choose between them. As discussed above, if Infl agrees with only the object, this triggers impoverishment and inverse voice. If instead Infl prefers the subject, or if it fails to decide between the two, then Infl will have distinct features from Voice and the direct form will be realized (and Infl will be realized as subject or portmanteau agreement). Oxford therefore proposes the following principles to decide which goal(s) to copy from, couched in terms of interaction/satisfaction.

(14) Probe P copies features from goal X if X meets the interaction condition of P and:

a. the features of X satisfy P, or
b. the features of X do not satisfy P, and thus the search needs to continue beyond X in the hope of satisfaction by a more distant goal Z.

Unpacking these principles for the current context of equidistance, the first requirement to be considered for copying is to meet the interaction conditions. So a goal that meets the interaction condition will be preferred over one that does not—this should not be surprising. Considering now the clause in (14b), if both potential goals meet the interaction conditions, but neither meets the satisfaction conditions, the features of both will be copied. Again, this makes sense. What *is* surprising based on the interaction/satisfaction model is the clause in (14a), which results in a preference to copy features from a goal that meets the probe's *satisfaction* conditions over one that does not.

Oxford needs satisfaction to bleed interaction in this way in order to account for cases in which the probe ignores the subject and agrees solely with the object. But by implicating satisfaction features in the algorithm for deciding which goals to copy features from, we have degraded, if not entirely removed, the distinction between interaction and satisfaction. We have given part of the job of determining which goal can be copied from (which is, by definition, meant to be the sole purview of interaction conditions) and given it to the satisfaction conditions (which, again by definition, should only dictate when the probe stops). This once again raises the question of whether such a distinction is necessary in the first place, or whether we can capture the underlying functions of selecting goals and halting the probe by other means. I argue that the proposed *u*F theory gives exactly the means to capture both functions through a unified set of features, with two further additions to the theory we've built to this point.

The first addition, which is adapted straightforwardly from the existing literature on Agree, is the possibility that features can be joined on a probe via either logical disjunction (Roversi, 2020) or conjunction (Scott, 2021). As we will see, this is also used implicitly in Oxford's analysis, and Deal (2024) admits the possibility to the interaction/satisfaction theory in her Footnote 10 (but, like me, does not find a use for it in the analysis of the PCC), and discusses the possibility in her other work (Deal, 2022). Let us refer to this general move as forming *logically joined* conditions on probes, or specifically as forming *disjunctive* or *conjunctive* conditions on probes. I define each as in (15).

(15) *Logically joined conditions on probes*

a. *Conjunctive conditions:* A goal G matches the conjunctive conditions $u\text{A} \wedge u\text{B}$ iff both A and B are found on G.

b. *Disjunctive conditions:* A goal G matches the disjunctive conditions $u\text{A} \vee u\text{B}$ iff A is found on G, B is found on G, or both A and B are found on G.

Note that both conditions are defined with respect to match between a probe and a goal. Recall: if match holds, the features of the goal are copied to the probe, and (if full match is achieved) the features of the probe are deactivated such that further cycles of Agree are halted.

The second addition to the theory is a principle of Best Match, which functions to allow a probe to decide which goal(s) to Agree with in case there is more than one that is equidistant to the probe (though nothing untoward happens if this principle applies when there is a single closest goal). This can be seen as a replacement for the conditions that Oxford uses to select goals under equidistance, and it also has various other conceptual antecedents in the literature (Coon and Bale, 2014; Oxford, 2019; Van Urk, 2015; Hammerly, 2021b, 2024)—though this particular formulation is novel to the current reply.

(16) *Best Match*

For two equidistant goals and a probe with features *u*A and/or *u*B...

a. If $u\text{A} \vee u\text{B}$, prefer most a goal that matches *both* $u\text{A} \wedge u\text{B}$; Prefer next-to-most a goal that matches $u\text{A} \veebar u\text{B}$; A goal that does not match either operand fails to match.

b. If $u\text{A} \wedge u\text{B}$ match only holds iff both operands match. A goal that does not match both operands fails to match.

Put another way, with both conjunctive and disjunctive conditions, Best Match will always prefer a goal that matches conjunctively. This makes sense: the conditions under which conjunctive statements match with the goal (iff both A and B are found on the goal) are a subset of the conditions under which disjunctive statements match (if A alone is found, B alone is found, or both A and B are found). Therefore, Best Match can be couched as a preference first for satisfying the conditions via conjunction in either case—the strongest possible statement consist with each logical connective. Then, if the probe is disjunctively specified, the probe tries selecting between goals via the weaker, but still valid, exclusive disjunction. Otherwise, the goals fail to match (leading no features to be copied) or match equally (leading the features of both to be copied).

All of this has been relatively abstract. Let us turn posthaste to considering and comparing Oxford's original interaction/satisfaction account with the proposed *u*F alternative.

5.2 Accounting for each case

In this section, I consider each type of inverse pattern described by Oxford (2022), showing that the *u*F theory can cover the same ground as the interaction/satisfaction account originally provided by Oxford. I reproduce the same basic tables as found in Oxford to show the patterns of inverse in each language. Each different pattern is labeled with C# or E#, matching Oxford's classification scheme of different types of Central and Eastern Algonquian languages (I refer the reader to the original paper for details). A few further notes: The rows represent subjects and the columns objects, so each cell is one particular alignment. Cells labeled "N" are realized with non-inverse voice (direct), while cells labeled "I" (also marked in grey) are realized with inverse. To review, the analytical goal in each case is to specify a probe on Infl that copies feature from the object alone in cases where the inverse arises, and agrees with either both the subject and object, or just the subject, in all other cases.

The first pattern in (17), which exists in both the Central and Eastern branches of the family, is a case where no inverse marking occurs. This can be accomplished with a probe with just φ within its interaction conditions (17a), or with just *u*φ (17b). Both will provide an equal match for both equidistant arguments (all are specified for φ), and therefore always copy the features of both the subject and object.

(17) *C1 & E1: No Inverse*

	1SG	2SG	1PL	2PL
ANIM	N	N	N	N
IMP	N	N	N	N
INAN	N	N	N	N

a. [INT:φ, SAT:–]

b. [*u*φ]

The pattern in (18) is described by inverse whenever the agent is an inanimate third person. We therefore need to prefer object agreement in these cases, but otherwise have an equal match. The key fact to leverage is that first, second, animate, and impersonal are all specified for [Anim], while inanimate lacks that feature. Both the probes in (18) will prefer or match best with the object just in case the subject is inanimate, resulting in inverse, otherwise both goals will be an equal match, resulting in direct.

(18) *C2: Inverse when agent is* INAN

	1SG	2SG	1PL	2PL
ANIM	N	N	N	N
IMP	N	N	N	N
INAN	I	I	I	I

a. [INT:ANIM, SAT:–]

b. [*u*Anim]

The pattern in (19) employs the same logic as the explanation for (18), but instead the decisive feature is [Pers], which only the third person animate and the local persons are specified for.

(19) *C3: Inverse when agent is* INAN *or* IMPERS

	1SG	2SG	1PL	2PL
ANIM	N	N	N	N
IMP	I	I	I	I
INAN	I	I	I	I

a. [INT:PERS, SAT:–]

b. [*u*Pers]

As in (19), the underlying pattern in (20) can in part be attributed to [Pers], which will prefer the local person object just in case the subject is inanimate or impersonal. However, when the subject is animate, an additional inverse case arises specifically with a 2PL object. Recall that for Oxford the system is set up such that, if two goals both match the interaction features (as all animate and local persons do), then the one that satisfies the probe is the sole goal that is copied from. Therefore, this can be captured by adding specific satisfaction conditions to the probe on Infl: [Addr] and [Pl]. While this is not directly discussed, these features must be understood as being specified conjunctively, so only a goal with both of these features will be preferred. This ensures that 1PL, which is specified for [Pl] but lacks [Addr], and 2SG, which is specified for [Addr] but not [Pl], will not be preferred by the probe. Only 2PL matches both, leading just the features of the object to be copied, which in turn triggers impoverishment and inverse.

(20) *C4: Inverse when agent is* INAN *or* IMPERS *or when* ANIM *acts on 2*PL

	1SG	2SG	1PL	2PL
ANIM	N	N	N	I
IMP	I	I	I	I
INAN	I	I	I	I

a. [INT:PERS, SAT:ADDR,PL]

b. [*u*Pers ∨ (*u*Addr ∧ *u*Pl)]

On the proposed *u*F model, we can capture this by conjunctively joining *u*Addr and *u*Pl, which are together *disjunctively* joined with *u*Pers. As already seen, when the agent is inanimate or impersonal, the object will always provide a better match since it matches at least one of the operands of the disjunctive statement (i.e. *u*Pers). With an animate subject, both the subject and object match the lefthand operand, as all goals are specified for *u*Pers. In the case of 1SG, 2SG, and 1PL, the subject and object are an equal match, since both only match via exclusive disjunction with *u*Pers. Therefore the features of both goals are copied and no inverse is triggered. However, 2PL is also able to match the conjunctive statement *u*Addr ∧ *u*Pl, and thus matches via conjunction with overall disjunctive statement on the probe. Best Match therefore dictates that only the features of 2PL should be copied to the probe, leading to inverse in this specific case.

The logic behind (21) is identical to what was discussed above with (20). However, instead of [Addr], the conjunctive statement (in either the interaction/satisfaction or *u*F type probe) is generalized to [Part], leading both 1PL and 2PL objects with animate subjects to be preferred by the probe, leading to inverse in these cases.

(21) *C5: Inverse when agent is* INAN *or* IMPERS *or when* ANIM *acts on 1*PL *or 2*PL

	1SG	2SG	1PL	2PL
ANIM	N	N	I	I
IMP	I	I	I	I
INAN	I	I	I	I

a. [INT:PERS, SAT:PART,PL]

b. [*u*Pers ∨ (*u*Part ∧ *u*Pl)]

Similarly, the logic behind (22) is the same as the previous two cases, but the satisfaction conditions and the righthand operand (as written) can be simplified to [Addr] rather than a conjunctive statement. This results in the two second persons (singular and plural) to trigger the inverse.

(22) *C6: Inverse when agent is* INAN *or* IMPERS *or when* ANIM *acts on 2*

	1SG	2SG	1PL	2PL
ANIM	N	I	N	I
IMP	I	I	I	I
INAN	I	I	I	I

a. [INT:PERS, SAT:ADDR]

b. [*u*Pers ∨ *u*Addr]

The last of the Central languages (which is also a pattern found in the Eastern branch) is when the inverse takes over in all forms. This can be captured by a probe that always prefers to copy features from the local person object. In the interaction/satisfaction model, this is a probe with SAT:PART. In the *u*F model, the probe can simply be specified for *u*Part (though nothing changes if other geometrically entailed features like [Anim], [Pers], and φ are conjunctively joined as well).

(23) *C7 & E7: Inverse in all 3 → 1/2 combinations*

	1SG	2SG	1PL	2PL
ANIM	I	I	I	I
IMP	I	I	I	I
INAN	I	I	I	I

a. [INT:PERS, SAT:PART]

b. [*u*Part]

Turning now to the patterns specific to the Eastern Algonquian languages, consider first the pattern in (24), where inverse arises just in case the object is 2PL. The logic is very much the same as we have already seen with, for example, (20). The difference is that the interaction conditions or the lefthand operand of the disjunctive statement (as written) is generalized to φ. As such, the subject and object are equally matched (or in the context of Oxford's account, will fail to match the satisfaction conditions) just in case the object is anything other than 2PL. With 2PL, the object will meet the satisfaction conditions or match the conjunctive statement within the righthand operand of the overall disjunctive statement, which will lead only the features of the object to be copied, resulting in impoverishment and inverse.

(24) *E2: Inverse when patient is 2*PL

	1SG	2SG	1PL	2PL
ANIM	N	N	N	I
IMP	N	N	N	I
INAN	N	N	N	I

a. [INT:φ, SAT:ADDR,PL]

b. [*u*φ ∨ (*u*Addr ∧ *u*Pl)]

For the case in (25), the logic is identical to that above, however the [Addr] feature in both probes is replaced with [Part], leading both plural local persons objects to trigger inverse.

(25) *E3: Inverse when patient is 1*PL *or 2*PL

	1SG	2SG	1PL	2PL
ANIM	N	N	I	I
IMP	N	N	I	I
INAN	N	N	I	I

a. [INT:φ, SAT:PART,PL]

b. [$u\varphi \vee (u\text{Part} \wedge u\text{Pl})$]

The last three cases add an additional layer of complexity, as there is no single natural class that accounts for all cases where inverse arises. Consider first the pattern in (26). The two natural classes are plural participants and second persons (this leaves out 1SG). Oxford's solution is to specify two sets of alternative satisfaction conditions. Again, while not explicitly framed this way, this amounts to having *disjunctive* satisfaction conditions. The object is preferred over the subject just in case it matches either set of satisfaction conditions (or both, as with 2PL).

In the *u*F model, this situation is handled by nesting an additional disjunctive statement. With two disjunctions, if any of the three operands match, then the features of the goal have the chance to be copied, as the conditions are met via exclusive disjunction. If any *two* operands match, then we have at least partially matched via conjunction. If all three operands match, then match is achieved entirely via conjunction. It follows from Best Match that a match of three or two operands will be preferred over matching just one, as matching via conjunction is preferred over exclusive disjunction. The 2SG, 1PL, and 2PL objects all match at least one operand of the nested statement $u\text{Addr} \vee (u\text{Part} \wedge u\text{Pl})$, and therefore matches via conjunction with the overall disjunctive statement. As the subject only matches via exclusive disjunction, the object is preferred and inverse is triggered. The 1SG object fails to match the nested portion of the conditions, and therefore only matches via exclusive disjunction overall (the same as the subject in all cases). This captures the lack of inverse in these cases.

(26) *E4: Inverse when patient is 2*SG*, 1*PL *or 2*PL

	1SG	2SG	1PL	2PL
ANIM	N	I	I	I
IMP	N	I	I	I
INAN	N	I	I	I

a. [INT:φ, SAT:ADDR, SAT:PART,PL]

b. [$u\varphi \vee (u\text{Addr} \vee (u\text{Part} \wedge u\text{Pl}))$]

The logic of the pattern in (27) is the same as detailed in (26), however the leftmost operand or the interaction conditions are specified to [Anim]. This leads inverse to spread across all cases where the subject is inanimate, as the subject will fail to match so the object will always be preferred.

(27) *E5: Inverse when patient is 2*SG*, 1*PL *or 2*PL *and* INAN → *1*SG

	1SG	2SG	1PL	2PL
ANIM	N	I	I	I
IMP	N	I	I	I
INAN	I	I	I	I

a. [INT:ANIM, SAT:ADDR, SAT:PART,PL]
b. [uAnim ∨ (uAddr ∨ (uPart ∧ uPl))]

Finally, the pattern in (28) also shows the same logic, but now the leftmost operand or the interaction condition is specified to [Pers], leading the inverse to spread across all forms with an impersonal or inanimate subject, as these goals will now fail to match, leading the object to always be preferred for copying.

(28) *E6: Inverse when patient is 2*SG*, 1*PL *or 2*PL *and* INAN/IMPERS → *1*SG

	1SG	2SG	1PL	2PL
ANIM	N	I	I	I
IMP	I	I	I	I
INAN	I	I	I	I

a. [INT:PERS, SAT:ADDR, SAT:PART,PL]
b. [uPers ∨ (uAddr ∨ (uPart ∧ uPl))]

This concludes the patterns identified by Oxford. Let us now move to some general discussion and comparisons between the two accounts.

6 TAKING STOCK

6.1 Selecting and copying

The main thrust of the argument in this reply is centred around deficiencies in the interaction conditions. Interaction conditions are not *neccessary* because a theory based in uF features, which does not appeal to the distinction, can cover the same empirical ground. This makes the interaction/satisfaction distinction redundant. Stronger still, I have argued that interaction conditions are not *sufficient* for determining which goal(s) to copy features from. To capture the patterns of Algonquian direct/inverse, the only way to resolve the selection of goals under equidistance within the model was to appeal to satisfaction conditions in addition to interaction conditions. Interaction conditions alone, which are meant to solely govern goal selection, are therefore insufficient to serve their intended function.

I do not see a clear way to salvage the model from the issue that arises from Oxford's analysis of Algonquian, save perhaps motivating a different approach entirely to agreement in the language family (e.g. one that does not lead to choosing between two equidistant goals). This would be an uphill battle, as Oxford's analysis is tied into the broader Algonquian morphosyntactic web in an intricate way, so messing with any one part of the system is certain to have downstream effects. We must therefore either reformulate the procedure for selecting between equidistant goals to only reference interaction conditions (it is not at all clear how this would work), or admit that there really is no principled distinction between the features of a probe that determine which goals to copy from and when to stop searching for goals to copy from. The latter is exactly what the uF theory presented here offers: a single set of features that can determine these functions.

6.2 Cyclic updates

Both the interaction/satisfaction model and the presented *u*F model share the general idea that a probe can be "updated" after each cycle of Agree. They also share the idea that the possibility of such updates is parameterized in some sense. For the interaction/satisfaction model, this is conceptualized via dynamic interaction, where certain features on goals bear a diacritic that can lead to updates to the interaction conditions of the probe. Some languages have goals with such features, while others do not. Furthermore, the particular features that dynamically interact can differ across languages. The idea presented for *u*F is instead whether the individual features of a probe are deactivated in an all-or-nothing fashion (features are only deactivated once all features match) or a some-is-enough fashion (when an individual feature matches, it gets deactivated). With the some-is-enough probe, each cycle of Agree where matching takes place winnows the active features on the probe, leaving an "active residue" (cf. Béjar and Rezac, 2009). The all-or-nothing probe was utilized in the same contexts that lacked dynamic interaction, and the some-is-enough the same contexts where dynamic interaction was utilized.

As instantiated, the two theories both share the idea that updates to a probe specifically restrict the possible goals on future cycles. However, there are a number of differences. First, the locus of parameterization in the *u*F theory is on the probe, while in the interaction/satisfaction theory the potential loci are on the goal. Conceptually, the probe is a more natural place for such a parameter to target, as this is what actually governs the agreement relation. Second, the deactivation parameter is more general, as it does not make reference to specific features, as in dynamic interaction. This makes the deactivation parameter simpler, and therefore all else equal it should be preferred. While neither of these arguments are necessarily decisive, on balance, they tilt the scales further in favour of the *u*F model.

6.3 Triggering and halting: A note on insatiable probes

On the *u*F model, any given cycle of Agree is triggered by the presence of *u*F features that have not yet been deactivated via match. The probe is halted when all *u*F features have been deactivated. On a close read of Deal's proposal (including not just the 2024 paper, but all papers that detail the model), there is no explicit statement of what actually triggers the operation Agree in the interaction/satisfaction model, beyond the basic description that "Agree is instead taken as an operation which is triggered by a lexical item when Merged into a structure" (Deal, 2022, p. 26). This cannot be the full story for two reasons. First, because it is not clear *how* a trigger for Agree is recognized—

the quote is just a statement of the fact that it must be recognized somehow. Second, because Merge is not always an immediate prerequisite for a given cycle of probing—a probe can undergo multiple cycles without re-involving Merge if there is more than one potential goal in its c-command domain. However, one can reasonably deduce from the model what the trigger must be: Satisfaction conditions that are not met. If meeting the satisfaction conditions is what halts the probe from engaging in further cycles of Agree, then it must be that having unmet conditions is what triggers Agree.

However, this is at odds with the proposed representation of an insatiable probe: A probe that *lacks* satisfaction conditions, and therefore copies features from every goal within its domain that meets its interaction conditions. Why is a probe that lacks satisfaction conditions *never* satisfied rather than *always* satisfied? This is more than a matter of interpretational ambiguity. As deduced in the previous paragraph, the existence of outstanding satisfaction conditions must be the driver of each cycle of Agree. Therefore, if there are no conditions to begin with, what actually drives Agree in the case of an insatiable probe? Either we must conclude that, in fact, this is not the correct representation of an insatiable probe within the theory (raising the question: what would the correct representation be?), or it must be something outside of having unmet satisfaction conditions that triggers Agree. Adding another principle for triggering Agree (especially if it is one that only applies in this specific case) will only make the theory more complicated, and therefore dispreferred, relative to the *u*F theory proposed here.

In the *u*F account, insatiable probes are instead modelled by having an *overload* of uF features and an all-or-nothing deactivation setting. Conceptually, this makes sense. Such a probe appears insatiable because there is no one goal that can match all of its features at once, thus deactivating it. It insatiable not because it lacks features, but because it is loaded with triggers for Agree, greedily wanting to match with a combination of features that is difficult or impossible to find.

6.4 Etiology or teleology?

Deal's wider argument about the "etiological" origins of Agree—creating redundancies to facilitate communication—deserves careful scrutiny. Especially as it is put forward as an argument against the *u*F model. Deal (2024) argues the following:

> "*Because this [the interaction and satisfaction] theory grounds Agree in the ability to create redundancy, rather than in the need to remove what Preminger (2014) calls 'derivational time-bombs,' it does not predict derivation breakdown in cases where redundancy cannot be syntactically established.*" (p. 45)

> "*From this perspective, it might be further speculated that the operations Merge and Agree, between them, represent the two major functions for which human language is adapted: Merge facilitates recursive processes of cognition and thus higher-order thought, whereas Agree facilitates communication and thus collaboration and social organization.*" (p. 45)

There are numerous points to unpack.

First, given that the current theory follows Preminger (2014) in that *u*F features are not *time bombs*, the argument in the first quote block is moot. The current theory is equally "grounded" in the idea that Agree makes redundancies—it just takes a different position on exactly how the representations and operations of syntax are set up to do so. Indeed, even the time-bomb theory is not necessarily boxed in to the position that interface crashes are the explanation for the existence of Agree in human language (though Chomsky's specific version, on some level, advances this claim). On this view, time bombs are simply a (flawed, likely incorrect) way of ensuring that Agree happens in any given derivation (i.e. a *trigger*, see the above discussion in §6.3), but they are not *necessarily* the "reason" for the existence of Agree in a diachronic or evolutionary sense. Therefore, it is important to distinguish these different levels of explanation: why Agree exists at all versus why Agree applies in a particular case (given that it exists).

More generally, Deal's etiological argument is based in teleological reasoning, or at the very least is ambiguous in a way that leaves the door open to such reasoning. This is most evident in the statement, "The operations Merge and Agree, between them, represent the *two major functions for which human language is adapted*". That is, Deal seems to claim that Agree arose as an adaptation in order to meet some functional communicative demands, making a functional pressure the etiological basis for Agree.

The first issue that arises from this type of argument is backwards causation (Mayr, 1988), where we assume the effect (communicative benefit) precedes or gives rise to the cause (a redundancy formation mechanism). That is, treating the system as goal-oriented such that Agree was "created" in order to fill some pre-identified functional need. This then requires us to commit to, for example, a vitalistic position (e.g. Bergson, 1907 [1959]), where we posit a special "life-force" that determines and orients the biological system towards goals. This only deepens the mystery of the etiological origins of Agree. Furthermore, one could easily imagine the same functional pressure (easing communication amidst a noisy world) giving rise to any other number of adaptations. For example, something physical like better hearing or sight, or something more squarely linguistic like an alternative algorithm for redundancy formation than the one we seem to have. Appealing to a broad functional pressure as a goal does not tell us anything about why this particular adaptation arose out of all the many possible ways one could imagine meeting that need.

This issue becomes even more profound when we consider the range of variation in agreement systems across human languages, from intricately rich to entirely absent. If we seek a goal-oriented evolutionary explanation that poses Agree as the solution to some fundamental functional pressure, it becomes tough to explain why some languages don't really have agreement at all (or have it in such a limited way). The most reasonable conclusion is that Agree is just one of the tools in the toolbox. Clearly, languages can lack agreement (and therefore presumably have a more limited role for Agree) and still adequately meet our communicative needs. Furthermore, Agree probably also plays a role well beyond redundancy formation qua agreement to have proliferated. Agree survives as a core part of our linguistic abilities despite "agreement" being unnecessary for any given language. All of this only makes it less likely that backwards causation, or any direct plea to functional pressures, is the right way to think about biolinguistic adaptation.

Where does this leave us? A simplified picture, which is essentially Neo-Darwinian, is that adaptations arise "by chance" (within genetic and other natural limits) and then are selected for (or not) based on their ability to meet the need posed by some functional or environmental pressure. This crucially distinguishes *generation* from *perpetuation*. Adaptations aren't "made for" anything specific, but they should be "used for" something(s) to proliferate in a population.

Returning out of the rabbit hole to the main point: It is difficult to evaluate Deal's evolutionary conjecture that the interaction/satisfaction model has a stronger etiological base than competitor models. I believe the question needs to be asked in a more precise way. For example, we need to ask whether a given model of Agree could have been generated based on our understanding of how new adaptations arise (e.g. chance mutations and the like), or whether communicative benefit is a sufficient selectional pressure for perpetuation of a given model of Agree. Answering these types of questions is not at all trivial, and much more work needs to be done.

As a first pass, we can use the standard outlined within the framework of Minimalist Inquiry by way of the Strong Minimalist Thesis (Chomsky, 2000, 2001): within empirical limits, we should posit the *most optimal* design for language to be usable. There are many ways to operationalize such a standard beyond Chomsky's view that this requires the specification of interface "legibility conditions", there are many specific "usability" conditions that one might consider, and there are always new understandings of the empirical landscape as data and patterns comes to light. As a result, "optimal design" is correctly thought of a relative notion. However, a safe starting point is parsimony: choose a simpler system over a more complex one.

Here is what is relevant: I showed that the empirical landscape used to motivate the interaction/satisfaction model, which assumes a more complex representation of an agreement probe with two sets of conditions, can be captured by a simpler *u*F model that does not make the same assumption, and only utilizes a single set of conditions. On the standard outlined in the previous paragraph, there is therefore no reason to prefer the interaction/satisfaction model over the one presented here. Stronger still, there is a basis to *disprefer* the more complex formulation in favour of the simpler one.

7 CONCLUSION

The current reply argued against the interaction/satisfaction model of Agree in favour of a souped-up version of a more traditional *u*F approach. To be sure, the underlying *concepts* behind interaction and satisfaction are still key. We must, in a principled manner, determine what goals a probe can copy from and when a probe stops its search. Additionally, we need to determine when Agree applies, as well as what features are copied from the goal to the probe. But, we do not need to encode representationally what can be derived algorithmically or factored out as a constant. This is at the heart of the argument in this reply: since we can capture the functions played by interaction and satisfaction without assuming a representational split, the split itself is not necessary. This does not diminish Deal's insight that any theory worth its salt must be able to capture both interaction and satisfaction in Agree. It simply derives it.

Acknowledgments Thanks to Will Oxford for detailed comments on an earlier draft of this reply. Special thanks to Reed Steiner and Anna Stacey for their participation and feedback in the context of LING 521 at UBC, where the ideas for this reply were developed and tested. All errors and claims are my own responsibility.

References

Béjar, Susana. 2003. Phi-syntax: A theory of agreement. Doctoral Dissertation, University of Toronto.

Béjar, Susana, and Milan Rezac. 2009. Cyclic agree. *Linguistic Inquiry* 40:35–73.

Bergson, Henri. 1907 [1959]. *L'évolution créatrice*. Paris: Presses Universitaires de France.

Chomsky, Noam. 2000. Minimalist inquiries: The framework. In *Step by step: Essays on minimalist syntax in honor of Howard Lasnik*, ed. Roger Martin, David Michaels, and Juan Uriagereka, 89–156. MIT Press.

Chomsky, Noam. 2001. Derivation by phase. In *Ken Hale: A life in linguistics*, ed. Michael Kenstowicz, 1–52. MIT Press.

Coon, Jessica, and Alan Bale. 2014. The interaction of person and number in Mi'gmaq. *Nordlyd* 41:85–101.

Cowper, Elizabeth, and Daniel Currie Hall. 2019. Scope variation in contrastive hierarchies of morphosyntactic features. In *Variable properties in language: Their nature and acquisition*, ed. David Lightfoot and Jon Havenhill, 27–41. Georgetown University Press.

Deal, Amy Rose. 2015. Interaction and satisfaction in φ-agreement. In *Proceedings of NELS 45*, ed. Thuy Bui and Deniz Ozyildiz. Amherst, MA: GLSA.

Deal, Amy Rose. 2022. Current models of agree. Ms., University of California, Berkeley.

Deal, Amy Rose. 2024. Interaction, satisfaction, and the PCC. *Linguistic Inquiry* 55:39–94.

Hammerly, Christopher. 2019. Limiting gender. In *Gender and noun classification*, ed. Éric Mathieu, Myriam Dali, and Gita Zareikar. Oxford University Press.

Hammerly, Christopher. 2020. Person-based prominence in Ojibwe. Doctoral Dissertation, University of Massachusetts Amherst.

Hammerly, Christopher. 2021a. A set-based representation of person features: Consequences for AGREE. In *Proceedings of NELS 51*, ed. Alessa Farinella and Angelica Hill.

Hammerly, Christopher. 2021b. A verb-raising analysis of the Ojibwe VOS/VSO alternation: Lessons for feature copying and movement. Manuscript, University of Minnesota, Twin Cities.

Hammerly, Christopher. 2023. A set-based semantics for person, obviation, and animacy. *Language* 99:38–80.

Hammerly, Christopher. 2024. The relativized EPP. Ms., University of British Columbia.

Harbour, Daniel. 2016. *Impossible persons*. Cambridge, MA: MIT Press.

Harley, Heidi, and Elizabeth Ritter. 2002. Person and number in pronouns: A feature-geometric analysis. *Language* 78:482–526.

Haspelmath, Martin. 2004. Explaining the ditransitive person-role constraint: A usage-based approach. *Constructions* 2:1–71.

Keine, Stefan. 2020. *Probes and their horizons*. MIT Press.

Mayr, Ernst. 1988. *Toward a new philosophy of biology: Observations of an evolutionist*. Harvard University Press.

McGinnis, Martha. 2005. On markedness asymmetries in person and number. *Language* 81:699–718.

Nevins, Andrew. 2007. The representation of third person and its consequences for person-case effects. *Natural Language & Linguistic Theory* 25:273–313.

Oxford, Will. 2019. Inverse marking and multiple agree in Algonquin. *Natural Language & Linguistic Theory* 37:955–996.

Oxford, Will. 2022. Probe specification and agreement variation: Evidence from the algonquian inverse. *Linguistic Inquiry* 1–47.

Preminger, Omer. 2014. *Agreement and its failures*. Cambridge, MA: MIT Press.

Rizzi, Luigi. 1990. *Relativized minimality*. MIT Press.

Roversi, Giovanni. 2020. How to satisfy probes: person/number hierarchy effects in Äiwoo. In *NELS 50*, ed. Mariam Asatryan, Yixiao Song, and Ayana Whitmal, volume 3, 99–112. Amherst: University of Massachusetts, Graduate Linguistics Students Association.

Royer, Justin, and Amy Rose Deal. 2023. Mayan animacy hierarchy effects and the dynamics of agree. Ms., University of California, Berkeley.

Scott, Tessa. 2021. Formalizing two types of mixed a/a′ movement. Ms., University of California, Berkeley.

Van Urk, Coppe. 2015. A uniform syntax for phrasal movement: A case study of Dinka Bor. Doctoral Dissertation, MIT.