Improper Movement in Phases

Since Chomsky(1973), the formal mechanism explaining improper movement (IM)(=1-2), has been representational, global and configurational (May 1979: Condition C approach, Fukui 1993: Chain Uniformity account).

(1) *Who seems that it is likely to win the race? (Saito, 2002)
[CP who5 [TP who4 seems[CP who3 that[TP it is likely[TP who2 to[VP who1 win the race]]]]]]

(2) *Who seems wins the race?
[CP who [TP who seems [CP who1 [TP who2 [VP who1 win the race]]]]]

This paper reanalyzes IM in terms of the phase-based derivational approach in Chomsky (2005:O(h)P(h)ases) and demonstrates that IM invariably results in a single-category feature-bundle violation of a bare output condition. The analysis we present empirically supports the C-to-T feature inheritance mechanism(C-T system) in OP, (and arguably deduced in Richards 2007).

[Two types of derivation]: In pre-OP systems, C and T are independent lexical entries. This system generates (3) as in (4): "who" is first attracted to [Spec,TP] by T, then "who" is attracted to [Spec,CP] by C. That is, T works as an independent probe as does C and so A-movement feeds A'-movement.

(3) Who bought the book?
(4) [CP who3[Case][Phi][Q] [TP who2[Case][Phi][Q] [VP who1 bought the book]]]

In OP, in contrast, T lacks inherent [phi], but inherits these features in the narrow syntax from C. Consequently T cannot function as a probe until C is introduced. In addition, C and T (the latter bearing inherited [phi]) separately attract elements from a single position simultaneously, so that (3) is generated as in (5).

(5) [CP who3[Q] [TP who2[Case][Phi] [VP who1[Case][Phi][Q] bought the book]]]

The crucial discrepancy between these two derivations is that the copies of "who" in [Spec,CP] and [Spec,TP] are related by movement in (4) but not in (5), in which there is no movement from [Spec,TP] to [Spec,CP]. [IM Deduced]: The IM data are, we claim, in fact correctly classified into two distinct types: Case on a moving element is valued after A'-movement in (1) and before A'-movement in (2). We demonstrate that both types of IM are appropriately excluded under the phase-based approach and it is especially the latter case that defines a clear empirical distinction between Pre-OP analyses, which in fact overgenerate (2), and OP analyses which yield the correct predictions.

(6) The derivation of (2) based on OP: *Who seems wins the race?
   a: [VP [who1[Q][Phi][Case]] wins the race]]
   b: [CP [who3[Q]] C [TP [who2[Phi][Case]] T [VP [who1[Q][Phi][Case]] [VP . . . ]] ]]
   c: [CP C[EF] [TP T[phi] seems[CP [who3[Q] [TP . . . ]]]]

The derivation of (2) is shown in (6) under OP. When C and T each attract "who" simultaneously (→6b), the features on "who" are separated: [Q] goes to [Spec,CP] and [phi]/[Case] goes to [Spec,TP]. In the matrix clause (→6c), only (the edge) "who" in embedded [Spec,CP] is visible to matrix C-T probing under PIC. Notice that "who" has only [Q], not [Phi] by virtue of feature separation, so that "who" is not an appropriate i.e. matching goal for the probing matrix T precisely because "who" lacks [phi]. In contrast, the matrix C can attract "who" but [phi] on matrix T is never valued, causing crash. This is a direct result of the C-T system, coupled with OP under which C and T separately attract different featural
subsets, simultaneously from the same launch site/Goal. The absence of \([\phi]\) on "who\(_3\)" makes it impossible for "who\(_3\)" to improperly move ("back") into an A-position: \([\text{Spec,TP}]\).

(7) The derivation of (2) based on pre-OP: *Who seems wins the race?
   a. \([vP < \text{who}\(_1\)[Q][\phi][\text{Case}] > \text{VP wins the race}]\]
   b. \([\text{CP}<\text{who}\(_3\)[Q][\phi][\text{Case}]>\text{C} \ [TP<\text{who}\(_2\)[Q][\phi][\text{Case}]>\text{T}\_\text{VP}<\text{who}\(_1\)[Q][\phi][\text{Case}]>\text{VP . . . }]]\]
   c. \([\text{CP}<\text{who}\(_5\)[Q][\phi]>\text{C}[\text{EF}][TP<\text{who}\(_4\)[Q][\phi]>\text{T}[u\phi]\text{seems}[\text{CP}<\text{who}\(_3\)[Q][\phi]>\text{TP . . . }]]\]

Under the pre-OP system, in contrast, feature separation does not occur, so that as in (7c), "who\(_3\)" in the embedded \([\text{Spec,CP}]\) bears lexically inherent and interpretable \([\phi]\) (and \([Q]\)). \([\phi]\) on "who\(_3\)" therefore can and does value \([u\phi]\) on the matrix T probe, then C attracts "who\(_4\)". The derivation converges. That is, pre-OP in fact overgenerates (2), while OP successfully excludes it. In addition, the other type of IM (1) is excluded straightforwardly: when the derivation reaches the embedded CP in (1), the transferred TP includes unvalued \([u\text{Case}]\) on "who\(_1/2\)" causing crash. The proposed analysis argues that IM is excluded by the independently motivated OP-analysis and without appealing to long-distance i.e. inter-phasal dependencies (e.g. binding relations or chains), so that computational complexity can be reduced in IM derivations to single-category featural-illegitimacy at the interface.

(8) a. John, likes himself, b. Himself, John, likes __. (Barss, 1986) (9) *He, likes John,

[Further support]: Moreover, (8b), empirically problematic for a Condition C account, is not problematic under the feature-based analysis. Also, our analysis does not unify the ungrammaticality of e.g. (1-2) with the interpretation in (9). This is the correct result as the anomaly of these data are in fact perceived/judged quite differently, and by hypothesis do not together constitute a natural class of violations (cf. Chomsky 1965, Epstein 1990).

(10) a. Who did you meet t? b. *Who t seems \([t \text{ wins the race}]\)?

Under the chain uniformity account, in addition, the distinction between an A'-A chain (→10a) and an A'-A-A'-A chain (→10b) is made by stipulating that the former is uniform but the latter is not, although, in fact neither is uniform since each contains both A and A’ positions. This is not problematic for us. The C-T system yields further consequences. Recall in (6), there is no movement from \([\text{Spec,TP}]\) to \([\text{Spec,CP}]\) since C-T independently attract elements from a single position simultaneously. This system predicts that C can only extract elements lower than the TP-domain. This in fact forces or explains Rizzi’s (1990) analysis of TP-adjuncts: a reason adverbial "why" is directly inserted into \([\text{Spec,CP}]\). TP-adjuncts are too high to be extracted i.e. probed by C, in the C-T system. Thus, External Merge of “why” into \([\text{Spec,CP}]\) is explained. Moreover, the OP-system split of a wh-phrase into \([Q]\) in \([\text{Spec,CP}]\) and \([\text{Case}]/[\phi]\) in \([\text{Spec,TP}]\) implies that non-branching lexical wh-phrases are in fact composed of two distinct morphological feature sets: WH on the one-hand and an indefinite QP "something" exactly as Chomsky (1964) proposed. Notice in addition that the parallelism requirement in (11) independently supports this morphological view.

(11) a. John bought something, but I don't know what John bought what.
   (John bought something, but I don't know what [John bought wh+something])
   (John bought a book, but I don't know what [John bought wh+something].)
   (John bought a book, but I don't know which book [John bought wh+book].)

[Conclusion]: The proposed analysis lends strong empirical support to OP in that IM deducibly crashes. The crash is localized to features of a single X\(_0\) and is neither global representational nor configurational. Moreover, the OP analysis, perhaps itself deducible, (Richards 2007) provides a derivational account of IM, which is empirically and conceptually
preferable to binding and chain-based analyses while entailing several disparate, but seemingly correct empirical consequences.

References
Chomsky, N. 2005. On phases. Ms. MIT.