The main claim of E&C’s comment on Lasnik’s article [1,2] is that generative grammar is built upon empirically weak (perhaps non-existent) foundations, and that generative grammarians aggressively resist experimental testing of their assumptions. Neither of these claims survives even brief scrutiny.

There is a long history of laboratory-based work that investigates foundational constructs of generative grammar. The operation merge, which produces recursive hierarchical structures, was already a focus of interest among psycholinguists 30 years ago [3]. Under the heading of the binding problem, the same issue is a leading problem in neuroscience research today [4]. The nature of long-distance dependencies, which some versions of generative grammar capture via movement operations, has been widely investigated using reaction-time methodologies since the 1960s [5]. The inconclusiveness of earlier results has led to progressive sharpening of the issues [6-9], particularly as concerns traces. Developmental considerations have been decisive in areas such as binding theory [10], and argument structure [11]. More recently, techniques from cognitive neuroscience have been added to the inventory of tools [12, 13]. This listing could be continued for pages.

Moreover, the list grows exponentially once we move beyond the methodological imperialism of E&C’s letter. Gathering of native speaker judgments is a trivially simple kind of experiment, one that makes it possible to obtain large numbers of highly robust empirical results in a short period of time, from a vast array of languages. Any good linguistics study involves carefully constructed materials, appropriate control items, and robust and replicable results. It is only because the technique is so easy and requires no more than a notebook that it is not usually described as an ‘experiment’. Note that when 4-year olds are involved, the same task calls for a quiet room, toys, and various clever ruses [14, 15], and then everybody agrees that it is an experiment. Outsiders would surely be puzzled by the attitude that seeks to deny the psychological relevance of easy, robust results, while insisting on other, far more subtle measures, such as 20ms differences in reaction times, or 1 second changes in how quickly babies get bored, or 2% changes in regional cerebral blood flow. The variability that one observes in native speaker judgments is real, but very small relative both to the agreement among speakers and relative to the variability in other measures (we have observed this repeatedly in our own studies). Furthermore, it is a truism in linguistics, widely acknowledged and taken into account, that acceptability ratings can vary for many reasons independent of grammaticality [16].
In language, as in any other area of inquiry, decisive evidence can – and does – come from a variety of sources, and it is hard to know in advance where the key evidence will come from. It is at best misleading for E and C to maintain that it is not clear how to turn the foundational assumptions of generative linguistics into testable hypotheses, since many researchers, whether with labs or with notebooks, have been doing so for decades [17].

References

[1] Edelman & Christiansen