We thank Tosh et al. (1) for their interest in our research (2) but note that their analyses do not undermine the main findings of our article. Their suggestion that polity population divided by polity area should be one of the social complexity dimensions raises a number of issues. What does this ratio mean at large spatial scales, where populations are concentrated in large urban centers and much of the territory is not heavily populated? How are societies distributed across this variable and why? For example, a small-scale “simple” society could have a very high population density if it has access to a rich resource base. Tosh et al. (1) do not provide sufficient information or context to meaningfully interpret their results.

The study by Chick (3), cited by Tosh et al. (1), was based on the Standard Cross-Cultural Sample, which is heavily weighted toward low-complexity, small-scale societies. The second component that Chick (3) found captures the differences in mobility and mode of production between agricultural and foraging societies, rather than complexity of organization per se. Tosh et al. (1) also argue without evidence that the proportion of variance explained by principal component 2...
“would likely have been higher had Turchin et al.’s data also allowed us to reconstruct agricultural intensity and urbanization.” In fact, our variable “capital population” is a proxy for urbanization (i.e., more urbanized societies would have larger population centers). Agricultural intensity (if interpreted as agricultural productivity rather than just whether agriculture is practiced) is likely to be strongly correlated with the complexity variables that we included in our study (4).

Finally, Tosh et al. (1) question our use of multiple imputation to deal with missing data. The number of counterintuitive cases produced appears low, and we already demonstrated that our results were robust with respect to the multiple imputation procedure and the degree of missing data. These techniques can always be improved, but we strongly disagree with Tosh et al.’s suggestion that it would be “wiser for researchers to estimate intervals manually” (1). A major guiding principle of the Seshat project is to reflect not only what is known but also what is unknown, by which we mean that experts do not have sufficient information to enter a value. Entering values manually based on guesses risks introducing unexplored assumptions and serious biases. Our paper is the product of a deep engagement between scientists and humanities scholars across a number of disciplines (5, 6). We aim to capture information across large expanses of space and time in a systematic but thoughtful way that is sensitive to the challenges of the historical and archaeological records.

We fully recognize that in a study of this magnitude, there will inevitably be inconsistencies, disagreements, and the need to modify or update the databank as new information becomes available. We have created a website (seshatdatabank.info/data/) to enable all to view our data and see the reasons behind coding decisions. We invite experts on past societies to use this web resource to make suggestions as to where data can be improved or added.