Cenozoic convergence between the Indian and Asian plates produced the archetypical continental collision zone comprising the Himalaya mountain belt and the Tibetan Plateau. How and where India–Asia convergence was accommodated after collision at or before 52 Ma remains a long-standing controversy. Since 52 Ma, the two plates have converged up to 3,600 ± 35 km, yet the upper crustal shortening documented from the geological record of Asia and the Himalaya is up to approximately 2,350 km less. Here we show that the discrepancy between the convergence and the shortening can be explained by subduction of highly extended continental and oceanic Indian lithosphere within the Himalaya between approximately 50 and 25 Ma. Paleomagnetic data show that this extended continental and oceanic “Greater India” promontory resulted from 2,675 ± 700 km of North–South extension between 120 and 70 Ma, accommodated between the Tibetan Himalaya and cratonic India. We suggest that the approximately 50 Ma “India”–Asia collision was a collision of a Tibetan-Himalayan microcontinent with Asia, followed by subduction of the largely oceanic Greater India Basin along a subduction zone at the location of the Greater Himalaya. The “hard” India–Asia collision with thicker and contiguous Indian continental lithosphere occurred around 25–20 Ma. This hard collision is coincident with far-field deformation in central Asia and rapid exhumation of Greater Himalaya crystalline rocks, and may be linked to intensification of the Asian monsoon system. This two-stage collision between India and Asia is also reflected in the deep mantle remnants of subduction imaged with seismic tomography.

* Corresponding author. E-mail: d.j.vanhinsbergen@uu.nl