Human action is built by actively and simultaneously combining materials with intrinsically different properties into situated contextual configurations where they can mutually elaborate each other to create a whole that is both different from, and greater than, any of its constitutive parts. These resources include many different kinds of lexical and syntactic structures, prosody, gesture, embodied participation frameworks, sequential organization, and different kinds of materials in the environment, including tools created by others that structure local perception. The simultaneous use of different kinds of resources to build single actions has a number of consequences. First, different actors can contribute different kinds of materials that are implicated in the construction of a single action. For example embodied visual displays by hearers operate simultaneously on emerging talk by a speaker so that both the utterance and the turn have intrinsic organization that is both multi-party and multimodal. One effect of this simultaneous, distributed heterogeneity is that frequently the organization of action cannot be easily equated with the activities of single individuals, such as the person speaking at the moment, or with phenomena within a single medium such as talk. Second, subsequent action is frequently built through systematic transformations of the different kinds of materials provided by a prior action. In this process some elements of the prior contextual configuration, such as the encompassing participation framework, may remain unchanged, while others undergo significant modification. A punctual perspective on action, in which separate actions discretely follow one another, thus becomes more complex when action is seen to emerge within an unfolding mosaic of disparate materials and time frames which make possible not only systematic change, but also more enduring frameworks that provide crucial continuity. Third, the distributed, compositional structure of action provides a framework for developing the skills of newcomers within structured collaborative action. Fourth, human tools differ from the tools of other animals in that, like actions in talk, they are built by combining unlike materials into a whole not found in any of the individual parts (for example using a stone, a piece of wood and leather thongs to make an ax). This same combinatorial heterogeneity sits at the heart of human action in interaction, including language use. It creates within the unfolding organization of situated activity itself the distinctive forms of transformative collaborative action in the world, including socially organized perceptual and cognitive structures and the mutual alignment of bodies to each other, which constitutes us as humans.