Guy Hoffman - Fall 2003

A Reading of
Jodie A. Baird and Dare A. Baldwin's
Making Sense of Human Behavior: Parsing and Intentional Inference

The Baird and Baldwin paper is primarily concerned with the riddle of action parsing, i.e. how we identify separate actions in an "objectively" continuous stream of motion. This problem is similar to the task of speech segmentation, given that the audio signal we perceive is in no way segmented into words or even into distinct sentences.

There is not much in terms of theoretical resolution in this paper, and its main point is in experimental data showing that our attention and perception is indeed guided by an immediate segmentation into actions that are meaningful to the actor's intentions. This was shown by comparing attention to tone placement within and between intentional actions. Similar results were produced in infants using habituation tests.

Baird and Baldwin suggest a two-tier mechanism enabling this phenomenon: on one hand we use our knowledge about activities to aid our parsing of actions (high-level mechanism), on the other hand, the authors suggest that we must have a lower-level skill for detecting structure in a behavior stream, enabling us to produce higher-level understanding of repeatedly observed actions.

I strongly agree with the high-level skill proposition, and I believe that much of our perception is driven by expectations and beliefs. I think it is not far-fetched to claim that the way we perceive the world is in a manner of "filling in the blanks" into our (innate?) expectations of how intentional action - and speech for that matter - are shaped.

However, I do not completely agree with the assumption of a statistical "blank-slate" domain-generic structure detection mechanism. Our activity environment is too varied and can be interpreted in too many ways to assume that a pure statistical learning mechanism would always converge in the same analysis of what intentional action is. I believe that we are innately wired to certain patterns of motion and grammars of actions that enable us to make sense of human behavior. We can still learn and are flexible enough to understand new actions, but we will not consider anything an intentional action just because it happens often enough in the same setting.

Naturally, a collaborative machine has to be able to parse a human collaborator's actions. Following the above, a pre-designed library of atomic actions and ways to achieve overarching goals should prove fruitful for this purpose. Perception data should be used to *classify* rather than construe what the currently observed intentional action is.