

Sentient companions predicted and modeled into existence: explaining the tulpa phenomenon

Kaj Sotala

Presented as a talk in the Toward a Science of Consciousness 2015 conference.

Within the last few years, several Internet communities have sprung up around so-called "tulpas". One website (tulpa.info) describes a tulpa as "an independent consciousness in the mind [...] an internal personality separate from your own, but just as human. They are sentient, meaning they have their own thoughts, consciousness, perceptions and feelings, and even their own memories." Members of these communities discuss ways of actively creating tulpas, and exchange experiences and techniques. Tulpas may interact with their hosts via e.g. auditory, visual, or tactile hallucinations, appearing as real people.

Although the existence of such beings may sound implausible, several related phenomena are known to exist. These include children's imaginary friends, dissociative personality disorder, and the "illusion of independent agency" (Taylor et al. 2003) where fiction writers report experiencing their characters as real.

I hypothesize that tulpas may arise from the combination of three factors. First, conscious thought acts as a "reality simulator", and imagining something is essentially the same process as perceiving it, with the sense data being generated from an internal model rather than from external input (Hesslow 2002, Metzinger 2004). Second, our brains have evolved to be capable of modeling other people and predicting their behavior, so as to facilitate social interaction. Third, according to the predictive coding model of the brain (Clark 2013), action and perception/prediction are closely linked: doing something involves us predicting that we will do it, after which the brain carries out backwards inference to find the actions that are needed to fulfill the prediction.

This allows for a tulpa-creation process in which the practitioner starts with imagining the kind of person they wish to create, and how that person would behave in different situations. The mental images produced by this process are picked up by the people-modeling modules of the brain, which might not be able to distinguish between imagined and perceived sense data, and they begin creating a model of the tulpa that is being imagined. Practitioners report their tulpas sometimes doing new and surprising things, which could be explained by the brain doing backwards inference to find possible "deep causes" of the tulpa's imagined behavior, whose other consequences are then simulated, causing the tulpa to act in ways unanticipated by the practitioner. Eventually, once the model and the practitioner's ability to imagine the tulpa become strong enough, there will be a selfsustaining feedback loop: the model of the tulpa creates new predictions of its behavior, which are experienced as happening, and these experiences are fed back into the model, giving rise to new predictions and behavior. By this point, the tulpa will be experienced as acting independently and separately from the "main" personality.

References

- Clark, A. (2013) Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, 36:3.
- Hesslow, G. (2002) Conscious thought as simulation of behaviour and perception. *Trends in Cognitive Sciences*, 6(6).
- Metzinger, T. (2004) *Being No-one: The self-model theory of subjectivity*.
- Taylor, M. & Hodges, S.D. & Kohányi, A. (2003) The illusion of independent agency: do adult fiction writers experience their characters as having minds of their own? *Imagination, Cognition, and Personality*, 22(4).