Intelligent Agents with Subjective Experience

Ana Pasztor (PASZTORA@CS.FIU.EDU)
School of Computer Science
Florida International University
University Park, Miami, Fl 33199

Number 7 of *Communications of ACM*, vol. 37, was dedicated to Intelligent Agents. All contributing authors agreed that there is a need for agents capable of learning. In (Rieken, 1994), Minsky notes: "The problem with this, of course, is that we still don't have suitable learning techniques. Lenat and I agree in the view that in order to learn as a person does, one will need to begin with a considerable body of built-in knowledge about a variety of effective ways to learn. The problem is we have not done enough research yet to know how to do this."

In this paper I present a conceptual model of human communication, which allows us to decompose the processes underlying our subjective experiences in general and our learning strategies in particular into *units* that can be readily changed, learned, and transferred from one intelligent agent (including humans) to another.

This model is largely based on Neuro-Linguistic Programming (NLP) and on Damasio's (1994) neuroscientific and psychological results and hypotheses.

At each moment in time, our subjective experience is manifested in our sensory modalities through images: visual, auditory, olfactory, gustatory, or kinesthetic images. According to Damasio (1994), "[i]f our experiences did not become images, however fleetingly, they would not be anything we could know. This is true even for those topographically organized representations that are not attended to in the clear light of consciousness, but are activated covertly."

Images are linked together through *anchoring* to form sequences which have consolidated into behavioral units or strategies (Dilts *et al.*, 1980). Anchors are acquired by experience, under the influence of our personal histories. According to Damasio (1994), they activate innate and acquired dispositions, which trigger 1. changes in body state called emotions, which then are signaled back to the brain as feelings, and 2. the release of selective neurotransmitters in sets of nuclei in the brain.

While the function of feelings is to bring our emotions into our consciousness, the result of the neurotransmitter responses is a "particular style and level of efficiency of cognitive process which" *accompanies* the changes in body state or emotions. It is manifested in qualities called *submodalities* in NLP, sub-components of each of the representational modalities, like (visual:) color, brightness, focus etc.; (auditory:) pitch, melody, rhythm etc.; (kinesthetic:) duration, speed, tempo etc.

In order to transfer subjective experience from one agent to another, or build believable agents, we need to be able to *recognize* its structure in ourselves and in others. Introspection with awareness of the representations which make up our subjective experiences can be learned and taught. Dennett (1988) seems to agree: "This putative grainlessness, I hypothesize, is nothing but a sort of functional invariability: it is close kin to what Pylyshyn (1980,1984) calls *cognitive impenetrability*... so what counts for an individual as the simple or atomic properties of experienced items is subject to variation with training ... Consider the results of 'educating' the palate of a wine-taster, or 'ear training' for musicians. What had been 'atomic' or 'unanalysable' becomes noticeably compound and describable...."

Emotions are to a great degree perceptible to external observers through behavioral elements such as body posture, accessing cues, gestures, eye movements, and language patterns. They allow us to recognize the building blocks of subjective experience mentioned above, and utilize them to build intelligent agents we trust and accept.

Even though NLP has been successful in eliciting people's subjective experiences by way of behavioral cues and transferring their elements from one person to another for therapeutic and other purposes, the model presented here needs thorough empirical testing. If validated, it could be used to build intelligent agents which not only *seem* to have subjective experiences, but actually have them.

References

Damasio, Antonio R. (1994). *Descartes' Error*. Grosset/Putnam.

Dennett, Daniel C. (1988). Quining qualia. In Marcel, A.J. & Bisiach, E. (Eds.), *Consciousness in Contemporary Science*, pp. 42-77. Oxford: Clarendon Press.

Dilts, R. & Grinder, J. & Bandler, R. & Bandler, L. & Delosier, J. (1980). *Neuro-Linguistic Programming: Volume I, The study of the Structure of Subjective Experience.*, Cupertino, Calif.: Meta Publications.

Rieken, D. (1994). A Conversation with Marvin Minsky about Agents, *Communications of the ACM*, 37(7), 23–29.