Hi Jackie

Here's my comments. Happy New Year!

Happing to keep working with you on this idea.

cheers,

cynthia

Hi Jackie,

It was a pleasure to have you in the class! I'm very happy to keep discussing your ideas with you on “civilized machines.’ It's definitely an intriguing concept.

I really like how you integrated many different perspectives into your framework for civilized machines. I’m going to give comments section by section. It’s long…(warning 😊)

**Intro:** I also often talk about the human-machine ecology. How we mutually depend on each other to maintain the overall balance. I don’t know if humans will every treat machines as “equals.” There are many grades of “full citizen” in human society. Even children and adolescents are not full citizens in the adult sense. Animals are accepted in society but not as full citizens. I think this is a different issue than whether or not people will treat machines (at some machines) as social entities. We treat children and animals as social entities, but we do not grant them full adult human status. I think social robots and humans will be able to engage in a new kind of relationship – something that takes on familiar dimensions as our human relationships and non-human relationships (i.e. animals). I like the way the social support literature frames relationships as the kinds of support that can be provided.

**Phenomenology:** I agree that this issue of having machines behave in a way that people can perceive and interpret its actions as socially intelligent is critical. This concept is closely tied to “believability” in animation --- the entity has to behave in a way that you ascribe a mind to it. Further, this mind must be familiar to us (not an alien mind). We need to be able to relate to it – we can relate to animals because we share sufficient common biology and evolutionary ancestry. We believe we share fundamental states like hunger, fatigue, pain, etc. We share aspects of or morphology with many species, and consider those that are more similar to us as more intelligent. We believe we can relate to an octopus, a bird, an elephant, a chimpanzee, a dolphin, etc even though we do not know EXACTLY what it is really like to be like a different species with different brains and bodies and life experiences.

**Childish machines:** This is problematic for me in the sense that human infants and children are profoundly social. They come out of the womb with pre-social abilities to encourage mom and dad to ascribe a human-like mind to them, even though their mind has not sufficiently developed to legitimately be treated as such. Evolutionary and developmentally this is critical because by having parents treat infants “as if” they were fully human with fully human minds --- then infants are able to LEARN now to be truly social and to develop and mature into full humans.

I think what Piaget is getting at is that it is believed that in early infancy, infants actually do not distinguish between them selves and the experiences they are having with the rest of the world. They assume what they experience is what everyone experiences. They do not distinguish that minds are separate. That happens later, when they develop Theory of other Minds – or mindreading abilities. Then they are able to take the perspective of others, realize that they are different from one’s own, and able to compare and reason about them. This develops gradually over the first 4 years of age. For instance, children pass the false belief task at 4 years of age (able to understand and reason that people might have different beliefs based on different experience).

But, yes, children often break social rules. These are set by culture and must be learned. This is why much of preschool education is more about socialization than classic ‘education’.

**From Politeness to Considerateness:** It is interesting to think about an autistic person. Are they polite or considerate? A high functioning autistic person can learn all the social rules. They can behave and act in a polite manner.

But they don’t viscerally understand others as we do. I would argue they do not have the ability to empathize, to take the other’s perspective, to be able to put them elves in the other’s position. There is no “feeling felt.” In short, they do not have a relational understanding of others by mapping self on to other and vice versa to understand commonality and differences. They must rely on
objective rules – they are much more machine like in this way. All this is because autistic people lack a theory of other minds. Do you believe that a machine must have a theory of other minds to be Considerate?

This is the difference between a polite smile and a smile from a shared feeling of joy.

Your example of the car seems to be getting at a machines ability to infer the intention from a person. Do you think that you can derive a set of rules to do this? Or do you think a machine has to be able to empathize with a human?

I think this is what you’re getting at when you say “Or can it think from a human’s perspective?”

What do you think of the relation of knowledge verses experience? Let’s say you can encode commonsense knowledge into a machine. Is that enough? Or does the machine need a lifetime of common experience to ground that out? To be able to relate through common experience? The idea of shared experience and relating is more closely tied to the notion of Common Ground rather than commonsense knowledge.

Civilized Machine: Your notion of engagement seems to be closely tied to the concept of adhering to expectations. When the behavior of something fits your mental model of what it ought to do, think, feel, perceive --- then the interaction is “natural”. However, when your expectations are violated, now a person must do a lot more cognitive processing to try to make sense of the other. Nass calls this “obeying social laws.” When these “laws” are violated (like perhaps interacting with an alien) then we must struggle to make sense of behavior that we have no intuition about. Now it is like being an autistic person who is struggling to understand the behavior of another human – which seems to be a mystery.

Exploring Human Needs: Good. Yes, there are biological needs and social needs.

Social Constraints: Good. Yes. Behaving in a socially appropriate way does not necessarily mean behaving in a “correct” way with respect to achieving some desired outcome. Again, high functioning autistic people often do “correct” things but socially inappropriate. They may respond to a greeting correctly by saying “hello” in acknowledgement, but not make eye contact or smile --- things that convey interest and pleasure in the company of the other. Social rules take into account the feelings of other people. They therefore depend on the other human.

Considereate Acts: To reword how I understand your ideas.

1. The machines behavior must be believable. It must convey having a mind that in consistent with human expectations of interacting with other social entities having minds (animals, people, etc.)
2. It must have a theory of other minds. It must be able to establish common ground with people
3. It must behave in sensible and socially appropriate manner.

Constructing Trust: Ok this is a big topic. It certainly builds on what you’ve already discussed so far. Many things factor into a humans sense of trust in another.

Being predictable is one factor – the person has to feel they have a good enough mental model of you to be able to have reasonable expectations for the other’s behavior. If this is not the case, then the other is an “alien” that is “opaque” to the observer.

Being dependable and mutually supporting is another factor – given the other is able to infer your needs and desires, it acts in a way to support them (it’s “got your back”).

Being caring can be another – that the other actively seeks out what your needs and desires are in order to help satisfy them. Mom’s are constantly caring for their baby – making sure the baby is always comfortable (fed, not too hot or cold, clean diaper, etc.) , engaged when awake, allowed to sleep when tired, etc. There is a big literature in “attachment” for mothers and infants and how the infant learns to trust it’s environment (that it’s a safe place where his needs will be met)… Hormonal changes happen with women when they become mothers (“the mommy brain”) to forge this bond with her infant – and if necessary to put the infant’s needs above her own.

Feeling and Truth: Yes. Humans acts based on beliefs not “truth.” Emotions, our experience, our senses all color what we believe to be true verses what is “true.” We in effect actively construct our reality. This is all we have to base our behavior upon.

Virtual Being: This sounds like the common ground argument. Good.

Mental Model: Yes, people apply their theory of other minds to understand the behavior of other animate things, including robots.

Constructing Trust with a Civilized Machine: Good. But having the machine learn social laws from experience vs having them programmed in seems important too. I think this is like our discussion in class about “authenticity” of emotions. People don’t feel that emotions can simply be “given” or “programmed in” alone. Life experience must play into it – the ability to ground out emotive states with life experience – and common experience with others.
Again, this is the “autism” argument. Will the machine be perceived as being genuinely considerate (being able to relate and empathize – ‘feeling felt”) or just polite (apply coded rules without “feeling felt”).

**Discussion**: In Japan they have this term “kansai”. It’s hard for a non-Japanese speaker to understand what this word means, but I think it captures aspects of what you mean by “civilized”.

**Truly Considerate**: Well, philosophers might get hung up on this. But as far as impact on people’s real life – if the robot always behaves as a “civilized machine” it will be a “civilized machine” with respect to the practicalities of every day life.

**Future Works**: Let’s keep talking about this. I think we can come up with a better “benchmark”

Good Work!

Cynthia

---

On Dec 15, 2006, at 7:54 AM, Jackie Chia-Hsun Lee wrote:

```
hi, cynthia
here is my final paper.
hri was a really fun class and it motivates me lots thoughts.
i hope we can keep working together to push this forward to my phd works.
thanks a lot
jackie lee
mit media lab
```

On Dec 14, 2006, at 1:22 PM, Cynthia Breazeal wrote:

```
Class,

if you still have materials to hand in please do so today. Grades are due tomorrow.

It was a pleasure and have a great holiday.

thanks

cynthia
<hri_cmachine-lee.pdf>
<hri_cmachine-lee.doc>
```