Answers for the Written Exam of the Contextual Area of my Qualifying Exam

Stefan Marti, August 31st, 2001

Question A

Last year, Ben Shneiderman started his CACM article with the following statement (http://www.cs.umd.edu/~ben/p63-shneidermanSept2000CACMf.pdf):

> *Human-human relationships are rarely a good model for designing effective user interfaces. Spoken language is effective for human-human interaction, but often has severe limitations when applied to human-computer interaction.*

How would you address this and similar comments that effective Human-Computer Interfaces should not try to mimic Human-Human Interaction?

Problem

Shneiderman's statement contains these two separate claims:

Mimicking human-human interaction is not effective in human-machine interaction.
Using spoken language is not effective in human-machine interaction.

The second one (speech) is obviously a subset of the first (human-human interaction).

Answer

I believe that both mimicking human-human interaction and using spoken language may be effective in human-machine interaction, *if the purpose of the human-machine interaction is to build a social relationship, or the machine's purpose is to behave socially.*

Origin of Shneiderman's approach

In his publications about user interfaces, Shneiderman clearly takes an anti-speech and anti-anthropomorphization stance. Shneiderman believes in direct manipulation where people interact with a machine as a tool. I will illustrate three of his main points concerning interface design (Shneiderman, 1997) with notes describing audio and video consoles:

1. **Continuous representation of the objects and actions of interest**. A mixing console consists of a large matrix of knobs and sliders (up to 10,000 elements and more, see Figure 1), where each object represents a certain feature—and only *one* feature. Furthermore, when the sound engineer presses a button or turns a knob, it is clear what its current value is. Additionally, the button will never be hidden behind hierarchical menus. To a trained engineer, almost all functionality of a console becomes clear by just looking at the desk.

2. **Physical actions or presses of labeled buttons**. Typically, on analog mixing consoles, there are no input elements except buttons, knobs, and levers. To manipulate the sound, the engineer physically presses buttons, turns knobs, and moves sliders. The elements that are manipulated are labeled explicitly. Even without the labels, their function is obvious through their location and physical orientation on the desk.
3. **Operations are immediate, visible, and reversible**. If the engineer makes an error, it is obvious which input element has to be manipulated to reverse the action.

Shneiderman describes the advantages of such interfaces for the user:

- Novices learn basic functionality fast, usually through demonstration by a more experienced user
- Experts work rapidly
- No need for error messages
- Immediate progress report: one can see immediately if one comes closer to the goal
- Less anxiety on the side of the user because actions can be undone
- User is confident because she feels in control and the system is predictable

Figure 1: *Typical analog audio mixing console. Its width is more than 3 meters. There is no alternative for these "monsters"; especially it is* not *possible to place all input elements on a standard computer screen without reducing the usability remarkably.*

As a music and video editor myself, I had the experience of learning and teaching other people exactly in the way described by Shneiderman.

In recent years, consoles were digitized, and started to include more and more novel input devices, most of them digital. Two negative examples that support Shneiderman's view:

- **Motorized faders**
- **Touch screens** replacing faders and knobs.

The idea behind audio consoles with **motorized faders** and knobs is that certain "snapshots" of the state of a console can be recorded, and played back later, which is useful. However, this means that knobs and faders can be manipulated not only by the operator, but also *by the machine itself*. For expert users, this feature is a Godsend, because it allows them to recall mix configurations from earlier sessions. For novices, however, this feature often leads to an unexpected outcome. For example, a user accidentally hits a memory replay button, and within the fraction of a second, all current settings of the several thousand knobs are lost—an incredibly traumatic experience for any novice user, especially if it happens during live audio mixing. It could destroy all trust in the machine and in the abilities of the user herself.

Another example that illustrates Shneiderman's points is the video-editing console I worked on during my time as a video editor for a national TV station. The user interface of the console consisted of a **touch screen** with a monochrome 10-inch monitor. The training for this system took several months. The reason

was that the operator had to gradually build up a mental image of what the system did, which state it was in, and what it was going to do next. The user interface gave almost no feedback, and important screens were hidden deep under hierarchical menus. Unfortunately, this is an example of how digital technology made the user interface worse.

Given these examples, one might understand why Shneiderman is so sure that human speech and human interaction style is out of question for certain human-machine interactions—it is indeed unthinkable to build a conversational speech interface to the above-described mixing console.

The new world

Yet, the above described situation and the respective user interfaces are only a small part of all possible human-machine interface situations. As Maes (1997) points out, Shneiderman seems to focus on a very specific domain, a professional user with a well-structured task domain, whereas Maes might focus on a different domain, e.g., a computer illiterate in an unstructured domain like the Web.

There are not only human-machine interface situations other than the ones described above, but also **our world is becoming more and more complex**. Certain human-machine interaction types cannot be approached efficiently with direct manipulation interfaces that Shneiderman describes. Just imagine an interface for the Web that tries to visualize all existing Web sites at the same time.

Our world has also changed in another important aspect: the **computational power available for user interfaces** has increased immensely. With this abundance of computational power, designers are able to think about alternatives to existing user interfaces and create tools that have more than simple utility, tools which allow us to be more human in the interaction with them: interfaces that are modeled after human-human interaction.

Because humans are naturally experts in human-human interaction, we are already familiar with such interaction styles. For example, human mankind always had slaves, servants, butlers, and assistants who served as "tools" for others, because the others did not have the skills to do a required job, or did not want to because it was straining. Therefore, humans are perfectly used to the notion of an autonomous entity that has some utility—the only difference is that such autonomous entities are artificial.

Assuming preferences for speech in human-machine interaction

Some researchers (e.g., Perzanowski 2000) "intuitively" *assume* that people prefer interacting with non-human entities in a human style. These researchers seem to have enough evidence to build complex multi-modal systems that are based on human speech and gesture. They simply extend human communication to incorporate humanoid robots. They argue that humans communicate with each other using certain "natural channels," such as talking and gesturing, and when interacting with machines, they prefer to do the same. In other words: If human-style communication channels are provided, and if the robotic agents look human, then humans will likely address these robots in human-like style.

Proof for preferences for speech in human-machine interaction?

Other researchers try to *prove* that humans do not care if the party they are communicating with is human or not, as long as they receive certain cues, upon which humans tend to apply social rules—even if they *know* that the other party is not human (Nass et al. 1993), and not a living entity (Miedaner in Hofstaedter, 1987). One of these cues happens to be human-sounding speech.

Nass et al. (1993) point out that ethnographic research and anecdotal evidences suggest that humans mimic human-human relationships in human-computer interaction. Before their experiments, the scientific community classified humans that behave that way as ignorant, or psychological or social dysfunctional. Before Nass et al.'s experiments, it was assumed that to provoke such social responses, one needs complex agents with animated faces, use of language, use of first-person references, etc. Their research however suggests that we need only minimal social cues to apply social rules of behavior to computers (Steuer 1995). Users know and believe that computers do not have "selves," yet they still behave as if they would. In addition, Dautenhahn claims that Nass et al.'s findings apply not only to computer, but also to any other agent, be it robotic or computational.

If human-machine interaction *can* be human-like, then speech is, as Nass et al. have shown, not only an option, but a must. If humans prefer to interact naturally with socially behaving entities, then it seems obvious that they will use spoken language.

On the other hand, Dautenhahn (1998) argues that human-machine interaction does not have to mimic human-human interaction, without rejecting the idea of agents. Dautenhahn is a strong proponent of autonomous agents, and focuses heavily on non-human entities in her research. She claims that to reach a *cognitive fit* between humans and their technological tools, one has to understand human perception, communication, and social and affective constraints. Dautenhahn explains that the human-tool interaction does not have to mimic nature and copy "natural" forms of interaction, but can be different. She hopes that some kind of "Interactive Intelligence" would emerge that is more than the sum of its parts, human plus tool: it would be a "dynamic spatio-temporal coupling between systems, embedded in a concrete social and cultural context." (p. 3) This kind of intelligence is different from intelligence in the classical sense, which is usually a sole property of the system itself.

The debate

Although Maes never emphasized spoken language as an interface modality to agents, the debate between Maes and Shneiderman is worth mentioning in this context, because it is in the early beginnings of this debate that Shneiderman started to express concerns about anthropomorphic agents and natural language interaction.

Nevertheless, the debate between Shneiderman and Maes is—as far as I am concerned—over. It seems pointless to me since there were viable solutions proposed to overcome the differences of the two approaches.

However, I would like to mention that I not only disagree with Shneiderman, but am also disappointed with the late Maes (Shneiderman and Maes 1997), who watered down the concept of "intelligent autonomous agent" to a point where it lost a lot of its original fascination, even rejecting the attribute "intelligent." True, it is not necessary to have human looking and speaking agents for, e.g., Web search related tasks. However, these two elements are essential if one enters the domain of, e.g., socially intelligent agents (SIA) (Dautenhahn) or socially intelligent autonomous robots (SIAR) (Breazeal).

The solutions

There are several approaches available that seems to combine the "best of both worlds." The concept of **Adjustable Autonomy** (AA) (Perzanowski 2000, Falcone 2000, Dorais 1998, Tambe 2001) is rather fuzzy, but demonstrates the idea of how to combine automation and direct user input. Yet, a clearer and more practical approach seems to be **Mixed-Initiative User Interfaces** (Horvitz 1999), which is based on AA principles.

Adjustable Autonomy in the User Interface

According to Perzanowski et al. (2000), AA is based on a slightly idealistic scenario that whenever humans interact, especially for task solving problems, they build teams, cooperate, start to assume roles, learn each other's strengths, and complement each other. Although these processes are complex, and still not yet explored completely, robotic agents have to be built to fit in this scheme. They must become team members, and they will adjust their autonomy as needs arise and change. The goal is to build a system that is autonomous: it knows enough about itself, the world around it, and what it has been doing, so that it can become a team player. The robots can act completely autonomously, but if necessary, they interact closely with their team members. The user or other agents, local or remote, can adjust the autonomy of these robots. The overall goal of Adjustable Autonomy is to create human-centered autonomous systems that enable users to interact with them at whatever level of control is most appropriate, whenever they choose, but minimizing the necessity for such interaction.

The point is that human-machine interaction can have human style, including spoken language, but depending on the situation, much lower level control should be available too. As long as the circumstances allow the robot to do its work autonomously, a user might feel comfortable interacting with it on relatively high communication level, which can be spoken language ("Work on project XY!"). However, in case of emergencies, the robot has to reduce its autonomy and get under closer control of the user or another agent, communicating on a much lower level ("Turn your camera twenty degrees left.")

Mixed-initiative User Interfaces

Another example where human-style interaction in the interface may be useful and efficient is mixed-initiative user interfaces (Horvitz, 1999). Although Horvitz does not discuss the need for spoken language directly, his list of design suggestions implies human-machine interaction that resembles human-style interaction:

He suggests that user interface agents have to be developed that:

1. add significant value over direct manipulation.
2. can deal with the uncertainty about the user's goals.
3. are aware of the user's attention (have a model of the user's attention), and don't interrupt at bad times (timing of services)
4. are aware of the costs/benefits of their actions, and take this into account.
5. engage in a dialog with the user to resolve uncertainties (but only if it is worth bothering the user!)
6. can be enabled and—more importantly—disabled easily.
7. try to minimize the costs of poor guesses: don't do stuff that could turn out very bad for the user
8. degrade gracefully if they are not sure about what is going on (anymore).
9. are ready to interact with a user, when she wants to, and even turns over unfinished work to the user, if she wants to.
10. behave socially correct, given their (social) role as a benevolent assistant.
11. remember what they, and the user, just did and said. "Shared short-term experiences," or memories of recent interactions (references to objects and goals), are important for a natural, comfortable discourse.
12. continue to learn from the interaction with the user and by looking over her shoulder. They should get better and better with time!

Important to note here is that most of these suggestions are agent capabilities that seem to be derived from typical human-style interaction. E.g., being able to deal with uncertainty is indeed a human characteristic, as well as being aware of the user's attention. Engaging in a dialog upon a problem is also human style interaction, as well as adjusting autonomy by handing over current work to a user (if she wishes), remembering what was said and done recently, and learning.

Finally, behaving in a socially correct way is a very important point that leads to the next topic: agents that behave socially or have social purpose.

Socially intelligent agents and robots

When the purpose of the machine is to engage the human in social interaction, considering spoken language and human-style interaction in the human-machine interface is appropriate, if not required, This is the most significant argument against Shneiderman's position.

A **Socially Intelligent Agent** (SIA) (Dautenhahn, 1998) is any kind of agent that shows human-style social intelligence. Applications for SIAs include games, virtual pets, and personal assistants that care for a single user. SIAs are appropriate if their function is primary social. They should be used when personality, character and personal relationships are desirable. Thus, SIAs' inherent expressiveness and believability must be in the right proportion to their intended functionality. For example, an SIA might not be appropriate if it requires too many resources (hardware, user's attention and cognitive load) to complete the task. There is a tradeoff between *efficiency* and *sociality*. Dautenhahn's SIA design guidelines, which are deduced from and strongly relate to our own human social behavior, include the following:

- Humans are **embodied agents**. Then, SIAs should be able to handle both objective and subjective time in human dialogues and in the way humans remember events and personal experiences.
- Humans are **active agents**, want to use their body and explore the environment. Then, the more degrees of freedom the user interface to an SIA has, the better (for the human).
- Humans are **individuals**, and they want to be treated as such, even if they have the same genotype. Then, developing *individuality* is important for SIAs: Imitation and social learning make agents more like us.
- Humans are **storytellers**. Creating and reconstructing stories is crucial for human understanding. SIAs have to be *good at telling and listening to stories*; that can be text, but also pictures, or non-verbal communication.
- Humans are **autobiographic agents** and **life-long learners**. They constantly learn and re-learn, re-write their autobiography. Then, SIAs could be helpful to re-construct autobiographical memories, strengthen social skills, and the self.
- Humans are **observers**. Human perception and cognition is subjective. Human behavior and motivations can only be understood in historical and cultural context. Along the same lines, agents have to adapt to cross-cultural differences.

What Dautenhahn suggests for agents of any kind is picked up by Breazeal (1999), who transfers the idea to the domain of robotic agents. Breazeal points out that today's robots are becoming more and more complex. At the same time, they interact more and more with lay people. Therefore, robots should be developed to interact naturally with untrained humans. I.e., they should be intuitive, efficient, and enjoyable.

Breazeal claims that a **Socially Intelligent Autonomous Robot** (SIAR) is doing exactly that. Its purpose is not only to transfer task-based information via intuitive communication channels, but also to address the emotional and interpersonal dimensions of social interaction with humans.

Note that we are not talking about traditional "robot appliances" that are designed to give the robot enough autonomy to carry out its task and still respond to commands of humans that oversee its performance, but an application that requires a more social form of human-robotic interface.

Breazeal describes four interface design issues for SIARs that are relevant in our context:

- **Human perception of SIARs**: How do people perceive SIARs?
- **Natural communication**: What channels of interaction are the most natural?

- **Affective impact**: How to interactions with SIARs impact people emotionally?
- **Social constraints**: what are the constraints in human-style social interaction?

Human perception of SIARs

According to Dennett (1987), our own and other people's behaviors are interpreted in terms of intentions, beliefs, and desires. Therefore, SIARs should be able to *convey intentionality*: They do not have to have them, but the user should be able to predict and explain the robot's behavior. Classical animators have perfected this art. However, Breazeal believes it is doubtful that superficial mechanisms of animation can be scaled to unconstrained social interactions between humans and SIARs.

Natural communication

Breazeal argues that speech and gesture are effective for task-based interaction, however, for more social interactions, perceiving the other's motivational state (beliefs, intents, wishes) is important. Such motivational state can be communicated through

- *Affective cues*: facial expressions, prosody, body posture
- *Social cues*: gaze direction, nods of head, raising eyebrows, etc.

It is also important to regulate the rate and content of information transferred (slow down, repeat). Most importantly, SIARs must not only send these cues, but also perceive them.

Affective impact of SIARs

However, there are dangers of using speech and human-like appearance. People anthropomorphize pets and empathize and bond with them, especially if they respond on a seemingly emotional level (e.g., a dog wagging its tail). SIARs may be in the same category as pets, but unfortunately, SIARs are not on the perfection level of a real pet yet. Therefore, one of the main challenges today is how to design SIARs that are not annoying or frustrating to users. Wrong expectations might be created by the SIAR's appearance. Good designs let the user interact with the robot at the exactly right level of sophistication.

As Norman points out (1994), the more we anthropomorphize agents, the more likely we create false hopes. E.g., speech recognition creates expectations of language understanding, pretending to have goals creates expectations of understanding human goals. Norman believes there are no moral problems as long as there are no false promises and no deception. In order to avoid false hopes, people need a "system image"—a basic understanding of how the robot works. However, complete transparency might have an undesired side effect. As Weizenbaum said quite some time ago, "*To explain is to explain away*. If something can be explained, it disappears. If there is a wondrous machine, and one eventually manages to explain its inner workings (in language sufficiently plain do induce understanding), then its magic crumbles away, revealed as a mere collection of procedures, each quite comprehensible" (Weizenbaum, 1966, p.23).

(Note that the "social intelligence" of SIAs and SIARs does not require the prerequisite of generic human intelligence. In other words, social intelligence is not based on generic human intelligence, but rather generic human intelligence stems from social intelligence. The *Social Intelligence Hypothesis* by Dautenhahn (1998) argues that today's generic human intelligence was derived from early social intelligence, which itself was necessary to deal with the increasingly complex social situation of early humans. At some point in the past, there seemed to have happened a transfer from social to non-social intelligence.)

Speech in the interface

Although I don't think that it is at the core of my contextual area, I would like to mention the *general problems of speech as an interface modality* in the HCI community, and that's where Shneiderman comes from.

Speech is slow, serial, transient, difficult to edit, and certainly not a general cure for all HCI problems. The ongoing controversy is on a rather emotional level, and some researchers seem to reject the idea of speech in the interface. It is not only Shneiderman who is very reserved towards speech, calling speech fun to use, but having too low of a bandwidth (which is true, except that the "fun" might be reduced remarkably if one does not speak English without an accent). E.g., Norman thinks that speech will not solve any problems that we have with the complexity of computers.

However, applied properly to the right situation, today's speech recognition has come a long way, and would be able to serve as an input modality, especially if the vocabulary and the syntax can be restrained, or even better, the machine trained to the user. The supporters just have to be clearer about the advantages and disadvantages of this interaction mode. For example, due to miniaturization of wireless communication devices where real estate is at a premium, speech interfaces could actually regain popularity with mobile devices.

Conclusion

In conclusion, I believe that I have found enough evidence to disagree with Shneiderman's statement that human-human relationships are not a good model for designing effective user interface, as well as that spoken language is not effective for human-machine interaction.

His statement may be applicable to certain classes of interactions, but not in the context of socially oriented robots and agents, which have the purpose to behave socially, or even try to build social relationships with their personal users. I hypothesize that this type of human-machine interaction will increase in the future, both because accurate modeling of human-human interaction is now possible, and because there is a natural human tendency to interact with any kind of agent in a human style, if the occasion is given.

Question B

"Beneath your desk you will find three iRX boards, a PIC programmer and a Chinese-English dictionary. **Create intelligent life**. Extra credit for obtaining legal guarantee of basic human rights for your creation."

Seriously, say you were given the above assignment and had the resources to pull it off. What **features would be most important** to implement such that others would see your creation as **worthy of being granted human-like rights**? What **features would not be necessary**? What **features** would be necessary before you, its **creator,** thought of your creation as worthy of those rights?

For the real question, assume you can get **whatever resources you need**, including (if you need it) technology that doesn't exist yet. What would you ask for? What features in your creature would be most important to achieve your goal, and why? What would be less important? What wouldn't be necessary?

Problem

What features are most important to create an entity that would be worthy of being granted **human-like rights,** both by **others** and **myself**, and what features are **not necessary**?

This question boils down to a set of similarly basic question: **What is humanness? What makes humans "human?" What is a soul?**

Answer

My proposal for features that are most relevant to obtain human-like rights are the followings (ranked in the order of their importance):

1. Features that lead to human-like **emotionality**
2. Features that lead to human-like **sociality** (including social behavior and intentionality)
3. Features that lead to human-like **intelligence** (in the traditional sense)
4. Features that lead to human-like **biological life-functions**

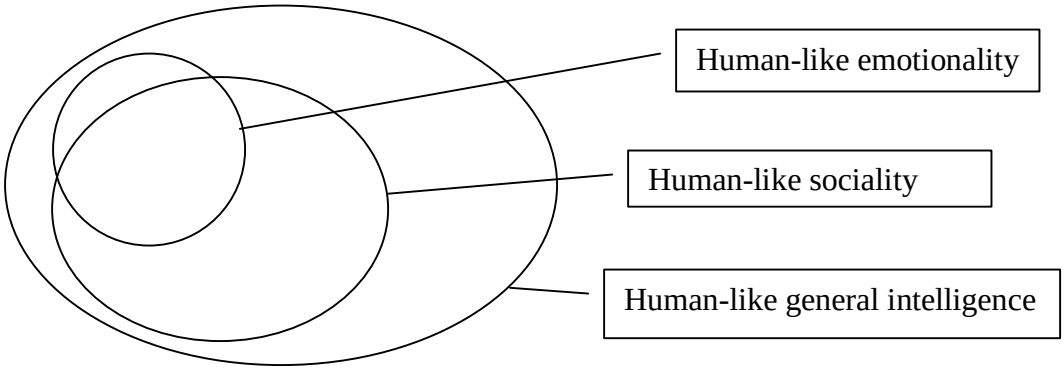

Figure 2: *Concentric feature categories*

I hypothesize that the most important features seem to be related to human-like **emotionality**. The next important class of features may be human-like **sociality**, followed by generic human-like **intelligence**. The least important class of features may be human-like **biological functions**.

The more feature categories a potential human-like entity can claim from the list above, the more likely it is guaranteed human-like rights.

Each category of features alone may not be enough to raise an artificial entity to a level where human-like rights are guaranteed. Thus, combinations of features will be necessary. I will try to identify what such combinations might be.

"Human-like emotionality" features seem to be almost enough to guarantee human-like rights

I hypothesize that human-like emotionality might be the main feature that would allow an artificial entity eventually to obtain human-like rights.

However, there is a striking example from Miedaner's wonderful science fiction, "The Soul of Martha, a Beast," in "The Mind's I" (Hofstaedter, 1981). Martha is a chimpanzee that can talk in simple sentences, directly generated in her brain and transmitted via a neural interface. She turns out to be a very believing, trusting, happy, child like character. She is intelligent in a broad sense, even in a human sense. Experts state that Martha is at least as intelligent as a human on the level of an imbecile. However, human-like intelligence does not guarantee human like treatment. The story's main character explains that, when such lab animals have outlived their usefulness, they get "eliminated." The researcher demonstrates this process publicly, and a poisoned candy kills Martha. During her (short) death struggle, her brain expresses her pain ("Hurt Martha Hurt Martha"), and then astonishment ("Why Why Why"), which appears to be absolutely heartbreaking to the audience and the reader of the essay.

This story leads the reader to the question: What is the difference between having a **mind (intellect)** and having a **soul (emotionality)?** Can one exist without the other? Is the degree of intellect a true indicator of degree of soul? Do retarded or senile people have "smaller souls" than normal people? Can we measure the soul through language? Is the Touring test a soul meter?

The essay demonstrates—on a very emotional level—the interesting idea that the **human mind could be linked very strongly to, or even defined through emotionality**. Intuitively, many people seem to agree that a soul is strongly linked to emotions. However, such an argument would merely transfer the problem of what is humanness to the question of how to disambiguate true emotions from simulated emotions.

Another example demonstrates that not even human-like appearance is necessary to evoke strong emotions and attribution of a soul. In the same book, Miedaner (in Hofstaedter, 1981) tells us the story of "The Soul of the Mark III Beast." The essay starts out with the assumption that biological life would be nothing more than a complex form of machinery. Therefore, human-built machines are just another life form. To illustrate that point, Miedaner asks the question if humans can relate emotionally to machines. Would it be possible to be bothered by breaking a machine, like killing an animal? He continues and explains that killing an animal is difficult because the animal resists death: it cries, struggles, or looks sad. Obviously, it is the person's mind that has a problem with killing. Then the author assumes a mechanical animal that exhibits animal like behavior, like sucking current from outlet, which equals eating, obstacle avoidance, which equals evasive behavior, and leaking lubricating fluid, which equals bleeding. Finally, he makes his point by saying that from the human emotional perspective, the destruction of such a machine would not be any different from killing a biological animal.

Miedaner finds that humans have no problems assuming "mechanical, metallic feelings." The question if one can kill an animal depends also a lot on the circumstances (drowning and ant in the sink, feeding live food to reptiles). People seem to sense that there is "soul-killing" going on in slaughterhouses, but they

don't want to be reminded of it. Humans seem to be animists to some degree, but **the souls we project into these objects are an image exists purely in our minds**. The author explains that we have a "storehouse" of empathy that we can tap into more or less easily; fleeting expressions, etc. can soften us.

But when does a body contain a soul? It appears **not to be a function of the inner state of the body,** but as a **function of our own ability to project.** This seems to be a rather behaviorist idea, since we ask nothing about the internal mechanisms. It also seems to be a strange kind of validation of the Turing test as a "soul detector."

Miedaner's essay is probably the most valuable article that I have read that addresses the problem how people might react to autonomous entities. In summary, the essay suggests that if there are enough cues for us to project a soul into an autonomous entity, it *will* have a soul, which in turn can raise it to a completely accepted being with all human privileges. The question remains, though, how few and which cues we actually need. I personally think that it might be even less than what Nass et al. (1993) suggest, as long as the context is appropriate.

"Human-like intelligence" might include the features of "human-like sociality"

Dautenhahn's (1998) *Social Intelligence Hypothesis* would support such a claim. Her hypothesis is that human intelligence originally evolved to solve social problems, and only later, it was extended to problems outside the social domain (mathematics, abstract thinking, logic, etc.) Thus, according to Dautenhahn, human generic intelligence originally came from social intelligence, which itself is necessary to deal with complex social situation of humans. More precisely, this hypothesis says that primate intelligence stems from adaptation to social complexity that occurred early in human race development. Dautenhahn suggests that it might have been a feedback loop, primate intelligence leading to increased brain size, which in turn enabled dealing with even more complex human societies.

However, there is a problem with the Social Intelligence Hypothesis. It can account for primate intelligence, but not for specific human intelligence. To solve this problem, Dautenhahn (2000) proposes the *Narrative Intelligence Hypothesis*: Stories seem to be the most efficient and natural way of human communication. Therefore, gossip and communication about third-party relationships differentiates us from other primates. We use our mental capacities to reason about other agents and social interactions.

"Human-like sociality" features alone are not enough

According to the above list of features that might lead to human-like rights, the next best indicator after human-like emotionality would be human-like sociality, or social behavior.

I have not found any relevant research that would support this hypothesis. However, I suggest experiments with robotic entities like Kismet (Breazeal, 1999), which might shed light on this question. Kismet is an example of a Socially Intelligent Autonomous Robot (SIAR). It is located in a deliberately benevolent environment and its sole tasks are to engage people in face-to-face social interaction, and to improve its social competence from these exchanges. The scenario is a robot child playing with a human caregiver.

Furthermore, Kismet is interesting because Breazeal describes synthetic emotions of a SIAR. She claims that emotions are important for social interaction. In constrained scenarios, the designer of an agent could profit from the anthropomorphization tendency that humans have naturally to achieve believable interactions. However, to participate in human-style interaction in unconstrained social scenarios, SIARs must be able to express and perceive emotions, which includes:

- **Synthetic emotions**
- **Empathetic learning mechanisms**
- **Affective reasoning capabilities**

"Generic human-like intelligence" features alone are definitively not enough

It is notable that up to now, research to create artificial intelligence in general, and artificial life specifically, has focused mainly on the part of intelligence that does not cover neither sociality nor emotionality: generic intelligence like logical reasoning, abstract mathematical calculation, etc. The reason may be that, e.g., mathematical intelligence is the easiest part to implement of all the above features.

However, we still seem to be far from this kind of human-like intelligence, and scientists are starting to ask the question why we don't have an "AI" yet (Stork 1997). In the book "HAL's legacy," he mentions that we might have met some visions of HAL, like speech, hardware, planning, and chess playing; but not in domains like language understanding and common sense.

In the same book, Minsky argues that although we seem to have good chess playing machines now, no one has ever tried to make a thinking machine and then teach it chess. He states that we have not progressed toward a truly intelligent machine. We only have some "dumb" specialists in restricted domains.

The closest we have to Minsky's suggestion may be CYC, a kind of intelligence that is based on common sense knowledge (Lenat in Stork 1997). The goal in building CYC was neither to understand how human minds work, nor to test some theory of intelligence, but just to build an artifact. Still, or perhaps because of that limitation, even CYC seems to be far from getting human-like rights. The question can be asked if CYC will ever get into the position to ask society for human-like rights.

Which features would *not* be necessary in order to obtain human-like rights?

Interestingly, most features of today's computers seem to be irrelevant, such as:

- Efficiency
- Consistency
- Perfection

It is somehow intuitive that attributes that we as humans lack will not be relevant features for a creature to get to a level where human-like rights might be granted.

However, there are reasons *not* to get rid of all non-human capabilities. Some researchers believe that the main advantage of technology including artificial agents (both robotic and computational) lies in the fact that its capabilities are complementary to the human capabilities (Billings 1997):

- Computers can calculate lots of data very fast
- Humans are flexible, creative, understand the world, and can reason with uncertainty and ambiguity

However, this also suggests that they are less human-like, and therefore may never receive human-like rights.

Besides the fact that scientist try to make artificial entities that are flexible and creative, what weakens this kind of thinking is that humans seem to prefer interacting in a human way, even with machines (c.f. the other question of this exam). This would mean that these artificial entities have to have an interface that is compatible with the human interface, including the social aspects of it. If we accept this "complementary thesis," then autonomous entities do **not** have be built like humans and do **not** have to be good at what humans are good at, but only need an interface that makes it possible to interact with them in human style.

As Dautenhahn mentions: today's robots are mainly associated with either machines (in a factory environment) or fictional characters (in movies). In the entertainment sector, television and movies, there are many examples for autonomous entities whose main problem is how to get human-like rights granted:

Star Trek Next Generation's Data, Star Trek Voyager's Doctor, A.I's. David. There are many more, but most of them are described in less sophisticated ways than the above mentioned.

Summary

In summary, here are the four main statements concerning categories of features that may allow an artificial life form to obtain human-like rights:

- *Human-like emotionality* features seem to be almost enough to guarantee human-like rights
- *Human-like sociality* features alone are not enough
- *Generic human-like intelligence* features alone are definitively not enough
- *Computer related features* are not necessary in order to obtain human-like rights

A combination of these features may grant an artificial entity with human-like rights.

References

Donald A. Norman (1994). *How Might People Interact with Agents*. Communications of the ACM 37 (7), July 1994, pp. 68-71.

Jonathan Steuer (1995). *Self vs. Other; Agent vs. Character; Anthropomorphism vs. Ethopoeia.* In *Vividness and Source of Evaluation as Determinants of Social Responses Toward Mediated Representations of Agency*, doctoral dissertation, Stanford U, advised by Nass and Reeves.

Lars Oestreicher, Helge Hüttenrauch, and Kerstin Severinsson-Eklund (1999). *Where are you going little robot? – Prospects of Human-Robot Interaction.* Position paper for the CHI '99 Basic Research Symposium.

Valentino Braitenberg (1984). *Vehicles: Experiments in Synthetic Psychology*. Cambridge MA: The MIT Press.

K. Bumby and Kerstin Dautenhahn (1999). *Investigating Children's Attitudes Towards Robots: A Case Study.* Proceedings of CT99, The Third International Cognitive Technology Conference, August, 1999, San Francisco CA.

Kerstin Dautenhahn (1998). *The Art of Designing Socially Intelligent Agents – Science, Fiction, and the Human in the Loop***.** Special Issue *Socially Intelligent Agents*, Applied Artificial Intelligence Journal, Vol. 12, 7-8, pp. 573-617.

Cynthia Breazeal (1999). *Robot in Society: Friend or Appliance?* In Agents99 Workshop on Emotion-Based Agent Architectures, Seattle, WA, pp. 18-26.

David Stork (ed.) (1997). *HAL's legacy: 2001's computer as dream and reality*. Cambridge MA: The MIT Press, chapters 1, 2, and 9.

Clifford Nass, Steuer, J., Tauber, E., and Reeder, H. (1993). *Anthropomorphism, Agency, & Ethopoeia: Computers as Social Actors*. Presented at INTERCHI '93; Conference of the ACM / SIGCHI and the IFIP; Amsterdam, Netherlands, April 1993.

Kerstin Dautenhahn (2000). Soc*ially Intelligent Agents and The Primate Social Brain - Towards a Science of Social Minds*. Proceedings of AAAI Fall Symposium *Socially Intelligent Agents - The Human in the Loop*, AAAI Press, Technical Report FS-00-04, pp. 35-51.

Kerstin Dautenhahn (1999). *Embodiment and Interaction in Socially Intelligent Life-Like Agents*. In C. L. Nehaniv (ed.) *Computation for Metaphors, Analogy and Agent*, Springer Lecture Notes in Artificial Intelligence, Volume 1562, New York, NY: Springer, pp. 102-142.

Robert D. Putnam (2000). *Bowling alone: The Collapse and Revival of American Community.* New York, NY: Simon and Schuster, selected chapters.

Robert D. Putnam (1995). *Bowling Alone: America's Declining Social Capital*. Journal of Democracy 6:1, January 1995, pp. 65-78.

Douglas R. Hofstadter and Daniel C. Dennett (1981). *The Mind's I: Fantasies and Reflections on Self and Soul*. New York, NY: Basic Books, chapters 4, 5, 8, 10, 11, 13, 18, 22.

Byron Reeves and Clifford Nass (1996). *The Media Equation*. Stanford, CA: Cambridge University Press, selected chapters.

Anne Foerst (1995). *The Courage to Doubt: How to Build Android Robots as a Theologian*. Talk, presented at Harvard Divinity School, November 27, 1995.

Joseph Weizenbaum (1976). *Computer power and human reason: From judgment to calculation*. San Francisco, CA: W.H. Freeman, pp. 1-16; 202-227; 258-280.

Joseph Weizenbaum (1966). *ELIZA: A Computer Program for the Study of Natural Language Communication Between Man and Machine*. Communications of the ACM 9(1):36-45.

Daniel C. Dennett (1987). *The Intentional Stance*. Cambridge, MA: The MIT Press.

Bill Joy (2000). *Why The Future Doesn't Need Us*. Wired Magazine 8.04.

Bruce Tognazzini (1994). *STARFIRE: A Vision of Future Computing* (video).

Erik Brynjolfsson and Michael Smith (2000). *The Great Equalizer? Customer Choice Behavior at Internet Shopbots*. Unpublished paper.

Dennis Perzanowski, A. Schultz, E. Marsh, and W. Adams (2000). *Two Ingredients for My Dinner with R2D2: Integration and Adjustable Autonomy*. Papers from the 2000 AAAI Spring Symposium Series, Menlo Park, CA: AAAI Press.

Rino Falcone and Cristiano Castelfranchi (2000). *Levels of Delegation and Levels of Adoption as the basis for Adjustable Autonomy*. Lecture Notes in Artificial Intelligence n°1792, pp. 285-296.

Michael Mogensen (2001). *Dependent Autonomy and Transparent Automatons?* In Lars Qvortrup (ed.) *Virtual Interaction: Interaction in/with Virtual Inhabited 3D Worlds*, New York, NY: Springer.

Dennis Peraznowski, William Adams, Alan Schultz, and Elaine Marsh (2000). *Towards Seamless Integration in a Multi-modal Interface.* Workshop on Interactive Robotics and Entertainment, Carnegie Mellon University: AAAI Press, pp. 3-9.

Eric Horvitz (1999). *Principles of Mixed-Initiative User Interfaces*. ACM CHI'99 Proceedings, pp. 159-166.

Ben Shneiderman (1997). *Direct Manipulation for Comprehensible, Predictable, and Controllable User Interfaces*. Proceedings of IUI97, International Conference on Intelligent User Interfaces, Orlando, FL, January 6-9, pp. 33-39.

Marc Mersiol, Ayda Saidane (2000). *A Tool to Support Function Allocation*. Proceedings of Safety and Usability Concerns in Aeronautics, SUCA 2000.

Gregory A. Dorais, R. Peter Bonasso, David Kortenkamp, Barney Pell, and Debra Schreckenghost (1998). *Adjustable Autonomy for Human-Centered Autonomous Systems on Mars*. Proceedings of the First International Conference of the Mars Society, Aug. 1998.

Alan Wexelblat and Pattie Maes (1997). *Issues for Software Agent UI*. Unpublished paper.

Ben Shneiderman and Pattie Maes (1997). *Direct manipulation vs. interface agents. Excerpts from debates at IUI 97 and CHI 97.* interactions, 4(6):42-61.

Dennis Perzanowski, A. Schultz, W. Adams, and E. Marsh (2000). *Using a Natural Language and Gesture Interface for Unmanned Vehicles*. In Unmanned Ground Vehicle Technology II, G.R. Gerhart, R.W. Gunderson, C.M. Shoemaker (eds.), Proceedings of the Society of Photo-Optical Instrumentation Engineers, vol. 4024, pp. 341-347.

Phoebe Sengers, Simon Penny, and Jeffrey Smith (2000). *Traces: Semi-Autonomous Avatars.* Unpublished paper.

Kerstin Dautenhahn (1999*). Robots as Social Actors: AURORA and the Case of Autism*. Proceedings of CT99, The Third International Cognitive Technology Conference, August 1999, San Francisco, CA, pp. 359-374.

Milind Tambe, David V. Pynadath, and Paul Scerri (2001). *Adjustable Autonomy: A Response.* Intelligent Agents VII Proceedings of the International workshop on Agents, Theories, Architectures and Languages.

Yasuo Kuniyoshi (1997). *Fusing autonomy and sociability in robots*. Proceedings of the first international conference on Autonomous agents, 1997, pp. 470-471.

Lenny Foner (1997). *What's an Agent, Anyway? A Sociological Case Study*. MIT Media Lab.

Charles E. Billings (1997). *Issues Concerning Human-Centered Intelligent Systems: What's "human-centered" and what's the problem?* Plenary talk at NSF Workshop on Human-Centered Systems: Information, Interactivity, And Intelligence (HCS), February 17-19, 1997, Crystal Gateway Marriott Hotel, Arlington, VA.

Brian Scassellati (2000). *Theory of Mind for a Humanoid Robot.* The first IEEE/RSJ International Conference on Humanoid Robotics, September 2000.

Cynthia Breazeal and Brian Scassellati (1999). *How to Build Robots that Make Friends and Influence People*. Presented at the 1999 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS-99), Kyongju, Korea.

Bruce Blumberg (1996). *Old Tricks, New Dogs: Ethology and Interactive Creatures.* Ph.D. thesis, MIT, chapters 1 and 2.

Justine Cassell and Hannes Vilhjálmsson (1999). *Fully Embodied Conversational Avatars: Making Communicative Behaviors Autonomous.* Autonomous Agents and Multi-Agent Systems 2(1), pp. 45-64.