On Being Supple:
In Search of Rigor without Rigidity in Meeting New Design and Evaluation Challenges for HCI Practitioners

Katherine Isbister
Polytechnic Institute of New York University
Six Metrotech Center, Brooklyn, NY 11201
isbister@poly.edu

Kristina Höök
Swedish Institute of Computer Science
Box 1263, 16429 Kista, Sweden
kia@sics.se

ABSTRACT
In this paper, we argue that HCI practitioners are facing new challenges in design and evaluation that can benefit from the establishment of commonly valued use qualities, with associated strategies for producing and rigorously evaluating work. We present a particular use quality—suppleness—as an example. We describe ways that use qualities can help shape design and evaluation process, and propose tactics for the CHI community to use to encourage the evolution of bodies of knowledge around use qualities.

Author Keywords
Suppleness, evaluation, prototyping, games, mobile, embodiment, use qualities.

ACM Classification Keywords
H.5.2 User interfaces

INTRODUCTION
As has been remarked by others (e.g. [4]) the last 10 years have presented CHI practitioners with a series of radically new challenges/opportunities:

• New contexts of use. Digital technologies are now commonplace in the home, in mobile contexts, and as part of ongoing social interaction and within complex social contexts.

• New inputs. Cheap, commercially available cameras and sensors finally in place with enough consumers mean we can take seriously the possibilities they provide to use for designing more nuanced and rich interactions.

• Great interest in new types of ‘tasks’. The growth of leisure and entertainment uses of technologies mean that we must stretch and evolve our design thinking to include support of process values such as fun, enjoyment, emotional engagement, and the like, including nonverbal interaction.

We believe all these factors make both design and evaluation profoundly more difficult for the following reasons:

• We can’t rely on re-using pre-existing interface metaphors and strategies because there are too many new variables of use.

• It’s hard to create designs without extensive prototyping and iteration. This means teams require new skill sets and more resources (including time).

• It’s hard to fully appreciate/investigate the results of this work (if they are not already a full-blown commercial product), so it’s hard for our research community to trade notes and build theory and knowledge about what we are doing.

• Evaluating the success of these interfaces requires new measures and ways to preserve context (e.g. social, environmental) otherwise tests just are not valid for the final use context.

We see this as one reason for the groundswell of dissatisfaction with traditional evaluation methods (e.g. [12]) to address the quality of our work in these new domains.

Yet our community has a strong empirical foundation. Despite a strong undercurrent of structural and theoretical critique, that is healthy in any case, we believe the vast majority of CHI attendees—researchers and practitioners—care deeply about grounding our work in rich observations of what does and doesn’t work, toward helping one another to design graceful and appropriate interfaces and interactions. We desire rigor in our practice. And we also desire the ability to extend and apply findings from individual projects to others—we don’t want to have to reinvent the process and practice as we go.

The authors believe a crucial first step toward meeting these new challenges is to cling closely to the rich end experience that we are trying to produce (a tactic that has become in-
fluent in recent HCI theory—e.g. [23]), then moving from the desired experience to an analysis of what makes it so and how it is crafted, toward extensible and applicable knowledge in broader contexts. If we ground ourselves in the fundamental experiential qualities we want to evoke for those who engage with our products, we believe it is possible to form a fresh understanding of how we can manage to shape these qualities, and how we can share our results with others and create systematic and helpful practices for evoking and evaluating such qualities.

We are inspired by the work of Jonas Löwgren [21, 22]. He has worked with colleagues to articulate a set of ‘use’ qualities of what he terms digital artefacts. These use qualities were shaped as a frame of reference for interaction designers to use in developing their practice. Löwgren and Stolterman explain that use qualities offer the value of ‘articulation of aims’, and are not meant to be a checklist or taxonomy but instead ‘proposed tools for questioning, elaboration, and making informed choices in thoughtful interaction design’ [p. 104, 22]. The idea is to set up touchpoints in dialog ‘with other designers and design theorists, as well as with design situations and the stakeholders involved in them.’ They note that the ‘main purpose of product quality articulation is to develop the ability to make judgments, which constitute a thoughtful approach to understanding the qualities of digital artefacts’ [p. 104, 22].

We find Löwgren’s concept helpful1 in pointing the way toward a new form of rigor that provides HCI practitioners with more flexibility. We believe use qualities can be taken up as pole stars to navigate by for designers and those who evaluate their work, as well as for end users and those who critique on their behalf (such as technology journalists).

We will illustrate what we mean in this paper, by defining and explicating a new use quality—suppleness2—that seems to come up in many recent user experiences that have the characteristics we outlined in this introduction (new contexts of use, new inputs, new types of task). We use this quality as a touchstone for discussing how our field could orient towards providing useful design knowledge to practitioners and still address the complexities of the new, experience-based, design situations. This paper therefore both defines what a supple design quality is, and through defining it, tries to shed some light on research methods and novel approaches for HCI-researchers to work in this field practically and with rigor.

Research Examples: Tales of Chasing Suppleness
Our interest in use qualities emerged from ongoing challenges we have faced in crafting research prototypes meant to engage people in rich and subtle ways. What follows here are brief descriptions of design issues that arose in the context of several projects, that help to illuminate the use quality we ended up defining using Löwgren’s approach.

Designing SenToy. SenToy is a plush toy with sensors inside the body [24], see Figure 1. By gesturing with SenToy, players control their avatar in a game. The six different gestures players can perform with SenToy were designed to resemble emotional, physical experiences. Shaking the doll angrily back and forth makes the avatar in the game angry. Quickly pulling the doll backwards, as if feeling surprised, makes the avatar surprised.

In our first studies of SenToy, it worked in many ways as we had hoped for. Users became absorbed by the game and their interaction through the doll. Their physical engagement showed how they both acted out their involvement through the doll, but also how they became very physical also in imitating the movements and expressions of their avatar. But unfortunately, in this first study, we discovered that the sensors and the wires inside SenToy were easily destroyed. The design team therefore decided to put a skeleton inside the doll to keep all the sensors in place. The skeleton was hard and only allowed for rigid, hard movements. It removed the floppy feeling of the doll and it became quite heavy. This changed the entire experience of the interaction. The sense of being one with the avatar went away. A more flexible skeleton would have allowed us to get back to the interaction we wanted, but the project could not afford creating yet another version of the system. In this case, what we’ve begun calling ‘suppleness’ was not privileged in the final research outcomes, yet we felt it as a loss in the end experience for users.

Designing eMoto. Another project that provided unexpected challenges in terms of the physical form factor of our prototype system was in the mobile messaging system eMoto, see Figure 1. [26]. The system made use of the stylus that comes with certain mobile phones, extending it with some sensors to allow users to express themselves physically. By gesturing with the pen, using pressure and movement, users changed the background to their text message to have colours, shapes and animations resembling their physical movements. Hard pressure and energetic movements with the stylus rendered a strong red colour with a large set of animated small, sharp-edged shapes moving in a jerky way. Less pressure and harmonic, wavy movements, rendered a calm blue, wavy background that slowly billowed back and forth. These messages could then be sent to other users to express various emotional content.

---

1 There are other HCI practitioners who have introduced terms to put names on hard-to-describe usability qualities—e.g. Green and Petre’s Cognitive Dimensions [11]—we found Löwgren’s approach and taxonomy particularly helpful, but the reader may find others that work well for their purposes.

2 We co-chaired a CHI workshop in 2007, using suppleness as an organizing theme. The contributions at that workshop also helped to shape the present conceptualization of the use quality ‘suppleness’.
While eMoto in many ways was a success, the actual shape of the extended stylus was a disappointment to both us and our users. Since we could not afford a proper design, the stylus became quite large in order to include battery, Bluetooth communication to the mobile, and sensors. Users therefore referred to it as a dildo and felt embarrassed to use it in public.

Another issue, both with SenToy and eMoto, was timing. The interaction between the physical device and the response on the screen had to be in perfect sync in order to make sense. It was not necessarily a matter of providing real-time feedback or responding instantly, but rather that different gestures entailed different experience that had to take different amounts of time. A negative, inwards gesture with eMoto users felt should take time to slowly change the background into what the desired darkly coloured, depressive expression. An energetic, upbeat movement needed a quicker response from the system.

With eMoto, we also had to experiment quite a lot with how to turn on the stylus and how to turn it off to save battery. Instead of being always on, users had to apply some initial pressure to “start” the pen. This also subtracted from their experience of being in sync with the system as this initial pressure forced them to express something physically that was not necessarily what they wanted to portray in their message. Again, limited resources in the research project did not allow us to fully experiment with the interaction to get it right, from a suppleness point of view.

In work on both of these projects, some end use quality eluded us, no doubt in part because we were not explicitly aiming for such an experiential quality that is hard to articulate in the first place. We felt a lack in the final experience in both cases, and as researchers we wanted to find a way to bring give a name to this subtle experiential quality we were seeking, and bring it to the center of our attention all along the way, during design and prototyping.

Designing Wriggle. Isbister is currently developing a gesture-based game that requires players to move their whole bodies in ways designed to provoke emotion [15]. Players wear knitted hats with Nintendo Wii controllers inserted into them (see Figure 1). The design team decided to put the controllers in hats, after doing some initial exploratory prototyping. We realized that expert players who hold the Wiimote in their hands tend to make minimal gestures to evoke game responses, once they realize this is possible with the accelerometer-based controllers. Mounting the controller on the player’s head forces her to make much broader movements—something that is more likely to induce emotion. One of our researchers knitted the hat prototypes, and from a purely visual standpoint, they have succeeded well in pre-testing with players. Players find them amusing and slightly embarrassing to wear, which seems to heighten their enjoyment and engagement with the game. However, the first version hats have slowly stretched during testing, leading them to sometimes allow a Wiimote to escape during vigorous head movement. This is obviously undesirable. Also, in the first prototype we require some button presses to advance the game, and players find it very awkward to access the buttons atop their heads. Creating such an unusual form factor for interaction has meant that we need to allow for more iteration and testing to really hone the interaction and make the experience feel the way we’d like. As with SenToy and eMoto, we feel that putting our finger on a use quality that could guide our iterations would help us to stay true to the end vision in our work.

Reflecting on the design of SenToy, eMoto and Wriggle

The physical and functional experience of using the systems discussed here is not something typically discussed in HCI. There is no systematic approach to addressing the felt aspects of their interactive use, though some younger CHI researchers have begun to try to get at these qualities (e.g. [8, 19, 20]).

The problems above are not solely to do with do with the actual shape or form of the physical devices we designed. Their meaning is only given through interacting with an application and therefore their use quality only arises over time. Løwgren names this its dynamic gestalt [22, 22]. His argument is that an interactive system cannot be studied merely by looking at or touching the object. We only understand it after interacting with it, sometimes during a very long time period. The meaning of one’s mobile phone does not come from its outer shape or the colour resolution of the screen. It comes from the interaction between organi-
sation of menus, the access to the address book, how easily one can open and play messages of various kinds, or whether one can playfully surf through the photos one has taken. It reveals its true nature, its gestalt, only after some time of interaction. It is in this light we can understand the problematic nature of designing eMoto, SenToy, and Wriggle. They did not reveal their true dynamic gestalt to us as designers or to the users until after being used for a while.

As we worked through how we might address these issues in our design process to better aim our efforts, we realized that there is not a lot of writing in the CHI community about these kinds of issues and explorations. We may share polished end examples of systems that have particular exemplary qualities, but we do not always link these to the articulation of the aims of the project and the iterations along the way that may help others to avoid dead ends and find valuable best practices to leverage. Conversely, systems that are successful in terms of achieving other aims may not be fully experienced by those who judge the work (likely they only read a paper) and thus crucial lapses in experiential power get lost to the research community and are uncritiqued.

We were excited to discover Löwgren’s notion of use qualities, and found the premise very helpful, but none of the qualities that he outlined were quite what we had in mind in the projects that we have mentioned. So we used our past experiences in our own research, and his framework, to formulate a use quality we could aim for in future work.

**DEFINING THE SUPPLE USE QUALITY**

According to the online Miriam Webster dictionary, one definition of supple is “easy and fluent without stiffness or awkwardness”. The metaphorical underpinnings of the term supple are crucial to grasping why this design value can be so tricky to evoke. It can be seen in two other definitions of the word: “capable of being bent or folded without creases, cracks, or breaks” and “able to perform bending or twisting movements with ease”. Suppleness is a use quality that depends upon:

**Subtle Social Signals**: Rich human communication and interpretation strategies (e.g. emotion, social ritual, nonverbal communication, kinaesthetic engagement). Supple interfaces are working beyond the cognitive/rational level of the standard GUI interface, both because of the use contexts they operate in (potentially emotionally charged, social and leisure situations), and because of the input modalities they may make use of (such as gesture, facial expression, biosensor data). A supple interface is one that enables and possibly enhances these subtle social signals. In a sense, a supple system is doing a sort of social/emotional ‘dance’ with the end user.

**Emergent Dynamics**: Taking into consideration subtle communication dynamics that require new thinking about system adaptability and feedback. For example, increased legibility of system moves to help users actively co-construct practice and meaning and push system boundaries in interesting ways. A supple system use fits smoothly and gracefully into my social and situational context as I interact.

**Moment-to-Moment Experience**: Privileging the quality of moment-to-moment experience both in terms of design and in terms of evaluation of success of design (e.g. a focus on engagement, pleasure, rapport). This requires flexibility in establishing the exchange between user and system.

We would place suppleness using Löwgren and Stolterman’s taxonomy of use qualities in the ‘immediate sensations of interacting’ cluster (versus user motivations, social outcomes, or structural features). The immediate use qualities outlined by Löwgren and Stolterman are pliability, fluency, immersion, and control/autonomy [p. 132, 22]. None fully gets at the subtle, social-seeming, emergent and dynamic quality we wanted to articulate.

**Some Supple Examples**

In the process of refining our definition of suppleness, we took up several well-known commercial products that we felt achieved the user experience we sought. We gave each product a close examination, working to articulate specific

---

Figure 2 **Left**: LocoRoco is a game designed for the Playstation Portable (PSP). **Middle**: The apple iPhone. **Right**: Nintendo’s Wii sports game, in this case tennis being played by four people at once.
design choices that lead to a supple experience, and also read commentary from the designers of the systems, and from journalists and players reporting on their own experiences with these systems.

LocoRoco
LocoRoco, see Figure 2, is a casual game designed for the PSP (Sony Playstation Portable) platform (see www.us.playstation.com/LocoRoco; for official trailer and website). The player uses a simple control scheme to navigate LocoRoco creatures through a game world in two dimensions. The LocoRoco can split into many, or join into one large creature. The bouncing of the LocoRoco is determined by an in-game physics engine, which creates very subtle and ever-changing motion on-screen. The LocoRoco sing as they move through the game world, and their voices change depending upon how many are present. The game became a very popular demonstration at the 2007 Game Developers Conference, and the designer, Tsutomo Kuono, was interviewed by several game magazines and online news sources (e.g., www.1up.com/do/feature?pager.offset=1&cId=3149607 and www.gameplanet.co.nz/mag.dyn/Features/2949.html). In the 1Up interview, this exchange takes place:

1UP: Please describe Loco Roco in a few words to our readers.

TK: [Laughs] I never thought of how to describe it like this. Hmm. It's a feeling comfortable game.

In the Gameplanet interview, Kuono says: "My favourite part is actions by multiple LocoRoco. Although the part has nothing much to do with the game play, this is what I most wanted to do. Watching LocoRoco gather to sing and pile up is really fun. Also their actions are not determined by certain motion data so you can never be bored watching them.

I also love the background music of the game. To keep the same music across the world, I created LocoRoco language myself and wrote the lyrics. LocoRoco sings as they roll, and the chorus is added / decreased depending on the number of LocoRoco so the song constantly changes. I believe focusing on such part that looks pointless is actually important.”

Kuono describes his process for getting buy-in for the project at Sony, and how he had to move from powerpoints to a working prototype that gave a better feel for the game mechanic and why it would be fun.

Playing LocoRoco is a supple experience because of the ever-changing, dynamic, and pleasant movement and singing of the LocoRoco creatures in response to the player’s actions. The game has a very satisfying moment-to-moment feel to it—as the designer said, it is a ‘feeling comfortable’ game.

iPhone
The Apple iPhone, see Figure 2, is another example of a supple experience in the mobile sector, this one not explicitly about entertainment and gaming. The iPhone allows users to use touch to manipulate the on-screen menus and applications, and re-orient the display if the iPhone is tilted. These novel interaction strategies provide a subtle but pervasive difference for the user in how it feels to interact with the device. This video review by David Pogue of the New York Times (gizmodo.com/gadgets/nytimes-pulphite/pogues-iphone-video-review-272554.php) pokes fun at the phenomenon that the iPhone became pre-launch, while highlighting the grace and subtlety of the interface itself. An expert on games and emotion (Nicole Lazzaro) discusses the unique emotional profile and relationship that forms with the iphone because of the soft and gentle gestures that are used in interacting with it (www.podtech.net/scobleshow/technology/1572/talking-emotional-software-design-in-games-and-iphone).

The iPhone is an example of suppleness achieved through the use of new sensor capacities that allow for a broader range of action from users. It is strong in the subtle social and moment-to-moment aspects of suppleness.

Wii sports
Wii sports, see Figure 2, was bundled with the Nintendo Wii platform. This set of games maps the physical movements of sports such as bowling and tennis, to on-screen game mechanics, so players can make use of moves they already know well from the real-world analog to quickly master and play the on-screen games.

Katsuya Eguchi, manager/producer of Software Development Group 2 at EAD, who is in charge of Wii Sports, said to IGN.com (wii.ign.com/articles/709/709218p1.html): “Initially, our goal was to create something very simple that anyone could just pick up and play, and because everyone knows sports we thought that would probably be the best setting…. Rather than complicate things by adding professional athletes or really realistic graphics, our main goal was to have the player experience in a very intuitive way the feeling of being in the game.”

Reviewers of the game agree: “In Rob’s house, Sports has already kindled the multiplayer fire with more speed and spice than a Concord passenger snorting wasabi. Over there, they're already comparing it to Samba de Amigo, because it’s so simple, approachable, and enjoyable. But more important than that from a Eurogamer's perspective, I suspect, is that it's a deceptive simplicity, with genuine subtlety lurking beneath the surface…. Tennis is my favourite. Flick the ball up and then swing to serve, with forehands and backhands played much the way you'd imagine, and height and spin available to those who master the gestures that command them. Knee-jerk criticism has focused on the fact the players move about on their own, but that's surprisingly unimportant in the long-term as you become obsessed with mastering your swing. Timing is essential.
Receive the ball on the right side of the baseline and you can place it wherever you like on the other side of the net—swing early to play it cross court, swing when it feels most natural to hit a safe return, or swing just as you feel the ball slipping away to push it down the line. Finding the sweet spots that keep the ball just within the tramlines is rewarding, and developing the ability to find them regularly is satisfying." (from eurogamer.net www.eurogamer.net/article.php?article_id=70561).

Wii Sports provides a dynamic moment-to-moment experience for players, that allows them to use their own movement to drive the interaction with one another. It is very strong in these aspects of suppleness.

**Guitar Hero and Rock Band**
Guitar Hero was a run-away hit game that combined special hardware with software—players operate the game using a life-size toy guitar with four buttons instead of strings. The original game was released in 2005, and has lead to two sequels. According to wikipedia: “The game's average review score by critics, according to Metacritic, was 91%. The consensus of most reviewers was the game had a great soundtrack (consisting of high-quality covers) and a guitar controller that was both fun and easy to use” (en.wikipedia.org/wiki/Guitar_Hero). In 2007 Harmonix released Rock Band, a game with a similar premise, in which up to four people can perform at once—lead guitar, bass guitar, drums, and microphone for a singer. The game has received high praise and at the time of writing this paper (according to Wikipedia) has sold over 3 million units (each including all the custom hardware).

Harmonix was able to create two successful products, argue press members who have covered them, in part because of a unique culture that includes immersion of the whole team in prototype iteration and experimentation. According to a Wired article about the design process for Rock Band: “Every Harmonix staffer is in a "band" that brings together people from different parts of the company, and they get together every week to rock out to the latest build, then pass their thoughts on to the designers.” Later in the article, the author adds: “This organic, iterative design process gets everyone in the company involved in tweaking all the different minutiae that go into accurately replicating the rock-star experience on plastic faux instruments.” (see http://www.wired.com/gaming/gamingreviews/magazine/15 -10/mf_harmonix_sb for the full interview).

Both guitar hero and rock band provide an astutely chosen range of movements for players that allow them to play out the social fantasy of being a rock star. Players must get the timing of their performance right to succeed, establishing a rich and emergent dynamic between player and game.

**MOVING FROM A USE QUALITY TO DESIGN AND EVALUATION STRATEGY**

Now that we have illustrated what we mean by suppleness, we hope to use it as a case study for how our community could take a use quality and make use of it to establish rigorous processes for evaluation, as well as sharing and advancing relevant knowledge for how to achieve the use quality. We believe those interested in establishing and exploring a given use quality should:

1. Develop a rich and helpful definition of the use quality, based where possible upon the insights of all relevant parties: designers, end users, evaluators, and technology journalists and other critics of the work. Where possible this definition should preserve the nuances and complexity of these experts’ understanding of what they are aiming for, and err on the side of multiplicity rather than brevity. Having touchstone examples of what is meant, that others can experience for themselves to aid their understanding, is very valuable here.

2. Glean information from those who succeed at producing the use quality, about their process, and look for commonalities in process across successful examples of the use quality.

3. Share research examples of attempts to achieve the use quality, with both summative evaluations as well as helpful process discussion that exposes unsuccessful paths toward the use quality and ‘lessons learned’. These examples may reveal important process steps that help to ensure the end quality; or they may have some prescriptions for ways to handle particular types of technologies and contexts of use that can be helpful to many designers.

4. Work to establish criteria for evaluating whether a use quality has been achieved, and hold those who report their work (in particular research work without established commercial success) to these criteria. Of course we must recognize that these evaluative criteria may look quite different than those we’ve seen in the past, as a community.

If we take the example of suppleness, we would say that the three factors described at the beginning of the paper (subtle social signals, emergent dynamics, and an emphasis on moment-to-moment experience) are the core of the definition of the use quality. We’ve grounded this definition with examples that can be readily obtained by readers so that they can experience for themselves what we mean. We included commentary by the designers, as well as by games and technology journalists, which helps to develop a rich language around this particular use quality that helps both designers and evaluators to ‘know one when they see one’.

We used the same sources (interviews and reporting) to gather some information about successful practices that led to the creation of supple experiences. Here are three recurring themes that we noticed:
Prototyping. Each of these experiences was something that was extensively prototyped and iterated. It would seem that designers in search of a supple experience must commit to doing extensive prototyping with both software and hardware, to a high degree of usable polish. The tuning and tweaking reported in games and other technology journalism is likely the tip of the iceberg in terms of creating truly supple final products. It’s important not to underestimate the rounds of revision necessary.

Evolution of the whole (versus piecemeal). Most of these systems involved thinking about hardware and software and social setting simultaneously (how will it feel to use this in this setting, on this hardware). In order to advance the design, the developers end up testing and revising based on high fidelity prototypes subjected to the rich context of the end situation. Game developers in particular are familiar with testing their projects in the genuine use context (usually testing products for themselves in their own environments). Rich context seems crucial to tuning supleness.

Emphasis on ‘feel’. In both the reviews and in the designers’ interview statements, the conversation turns again and again to how it feels to engage with these interfaces. It seems that a successful designer of a supple experience likely has (or evolves) a strong, holistic vision of the feel of the interface that can be shared with others on the development team, toward iteration and improvement. Sometimes a creator has a driving vision from the beginning that s/he carries through and uses to make decisions as the project develops (e.g. LocoRoco). Other times, great ideas evolve from tinkering in a bottom-up way with the materials at hand (e.g. Twitter). And sometimes, a strong initial concept is replaced by a stronger revision based upon early prototyping and testing. The key is to recognize the strength of the iteration (what makes it really work) and to preserve that integrity through the rest of the refinement process. Game designers have begun to establish some language for sharing their vision about feel [e.g. 27]; we believe use qualities are another tool for practitioners in communicating and iterating toward a shared vision.

If we circle back to our own research prototyping, in retrospect we could say that the end experience of our work would have been stronger had we set for ourselves a focal use quality such as supleness explicitly, and assumed that we’d need multiple rounds of prototyping to achieve it, with careful attention to the feel of the whole system (hardware included) as experienced by users in each iteration. Had we used suppleness as a guiding concept in this way, we could have better directed our efforts so that the end result was a more powerful experience. We could have focused some of the user-testing of our prototypes on measuring in some fashion, whether users experienced this supple quality.

THIS LEADS US TO EVALUATION...

If a use quality emerges in a dynamic gestalt, over a period of use, how can one dissect and analyze and evaluate the presence or absence of such a quality? How can one craft a partial prototype and query potential users about what it is like to engage that prototype? If the end use will be situated in a particular social context, how close does the design testing have to be to this context for the results to be useful and ecologically valid? Evaluation is a fundamental challenge of designing the kinds of systems we are now tackling in the CHI community. If one cannot readily dissect a given use quality into clear sub-criteria to be checked off a list, how can one construct a questionnaire to elicit from users whether that quality was present? And yet this work is being done all the time—to some degree, it has been invisible work that doesn’t have a clear publication path in the CHI community, as was debated at last year’s conference in the panel on usability as sometimes harmful [12].

We need evaluative practices that can preserve the richness of the user’s unique and perhaps not fully and neatly describable experience with our system, and practices that allow them to express things that they might not be able to easily convey with words, that we can still make use of in iterating and evaluating our designs. There are some emerging alternative evaluative practices in these areas, which could be taken up in tackling particular use qualities:

Bodily experiences. When it comes to studies of designs that are underway, there have been attempts to involve users’ own reflections and not only cognitively, but also bodily. For example our Sensual Evaluation Instrument (SEI) [14] and the work of Sengers, Boehner, and Kaye at Cornell University (e.g. [1,2,3,19]). Today, the phenomenological movement in HCI is also putting tangible interaction at core [6,7], with the argument that our human practices typically involve physical, bodily practices and that interaction should be building on that understanding. There may be ways to use these techniques to get at the bodily sensations of system suppleness (or other use qualities) in evaluations.

User reflection. Evaluation of use qualities may be more effective when we give users room to reflect upon their experience with a system in rich ways. This was done, for example, in the study of eMoto where the users helped interpret their own data [26]. The users in this study also recruited a close friend or partner (spectator) who provided input on how the system was understood and used. The study method was inspired by the Cultural Probes method [10], and the Experience Clip method [16] that evaluates systems by making two friends use it and document each others use. Users could be offered a working definition of suppleness or other use qualities and asked to reflect on whether the system achieved this quality in their experience of it.

The Repertory Grid Technique is an old method that has been revived in order to get at how users subjectively perceive different designs [8]. By repeatedly providing users with three different objects and asking them to name a quality that differentiate one from the other two, the evaluator can step by step arrive at which qualities are most
prominent. What is interesting is that the categories by which the objects are classified are not pre-determined. The method has been modified to work for the dynamic qualities of an interaction through asking users to compare three different video clips from their usage of some interactive system [20]. We may be able to find clusters of terms that emerge from this type of evaluation that confirm or disconfirm a user’s experience of a given use quality.

Gaver has also proposed that cultural commentators, such as literature critics, journalists, or documentary filmmakers, could be used to provide multi-layered understandings of how people perceive, interact with and reflect on new technologies [9]. Such commentators could be given information about use qualities we hope to achieve and reflect upon whether they have been achieved, in their commentary.

A work in progress. The methods described here aim to capture some of the hard to get at dynamic, emotional, physical, subjective qualities of the designs that one aims for with a use quality like suppleness. But there is not yet a large body of examples of the use of these methods, nor robust discussion of their powers and limitations.

At present, we would argue, designers going after these use qualities are largely reduced to the kind of intuitive and reflective practice described so well by Schön [25]. He argues that practitioners’ knowledge that arises from having to deal with specific experiences of specific cases/situations can be ‘demystified’ and reflected upon. Through such a reflective process, practitioners gain knowledge that can best be described as a repertoire of design solutions. This repertoire can then be used, together with more generic ‘scientific’ knowledge, to address a specific case.

Similarly, we believe that we can gather a design repertoire from good products that provide good grounds for a supple experience. Applying the obtained knowledge from such a design repertoire to the specific case at hand, is not and cannot be a mechanistic process. A supple experience requires fine-tuning of exactly how we physically interact with the system, the timing of the responses from the system to user action needs to find its delicate balance, and the interaction will not look the same between different design cases. Or as Schön puts it:

“In real-world practice, problems do not present themselves to practitioners as givens. They must be constructed from the materials of problematic situations that are puzzling, troubling, and uncertain. In order to convert a problematic situation to a problem, a practitioner must do a certain kind of work. He must make sense of an uncertain situation that initially makes no sense.” [25, page 40]

Schön talks about this process as attempts to frame the problem. A designer will try to frame a specific problem in different ways and then look at how the problem then talks back in order to decide both the problem its solution. As pointed out by Schön, there is knowledge to be obtained from the specific cases if we involve ourselves in “reflection-in-action”.

Orienting use qualities, like suppleness, will help designers to frame their design problems and reflect on their designs. And the specific design cases can become a design repertoire of embodied knowledge.

We can also set use qualities as experiential anchor points, and begin to shape evaluation practice around these anchor points. We can look for signs of emotion, of smooth practice, we can seek feedback from users about how they feel and if they enjoy our prototypes in the ways we anticipate (or in ways we did not expect), making use of some of the methods we outlined above and developing new ones that suit our needs. We believe a key challenge in HCI practice in the coming years is advancing methods and practices for creating honest and rigorous dialog with our users (and among ourselves) about what our target use qualities are, and whether they have truly been achieved.

THOUGHTS ON ACHIEVING SUPPLENCESS AND OTHER USE QUALITIES

In the design and evaluation process:
There are a few things we see as potentially helpful interventions in the design and evaluation process, to create ideal conditions for producing interfaces that embody subtle and holistic use qualities such as suppleness.

Deep knowledge of materials. We need to respect, and cultivate, deep knowledge among practitioners of the materials they are working with—there seems to be a misperception in our educational programs and research lab configurations/hiring philosophies that one can move as a researcher seamlessly between, for example, mobile application design and immersive 3D worlds or sensor-based interaction in games. The more hands-on experience one has with working with affordances of particular materials and contexts, the more likely one is to produce a supple (or pliable or fluid or immersive) design. Use qualities like suppleness emerge from the details of the moment-to-moment unfolding of the experience—something hard to understand without tinkering for a while. This means project teams in both research and industry should highly value team members with materials expertise and should seek to leverage this knowledge when moving into a new domain.

We believe there is room for much more collaboration with practitioners from professions like toy design, fine art, or Hollywood special effects and mechatronics (as in the collaboration between Stan Winston and Cynthia Breazeal—see [5]). Toy designers, for example, might have been able to talk to us about the importance of floppiness of a doll and thereby would have informed our design of SenToy. Right now, such practitioners are rarely found in our community. We must find ways to involve them, and to share with them what we’ve learned. We probably need to change the form of communication of results to make this happen. More hands-on experience of our prototypes is likely needed.
Rich iterative prototyping. We need to acknowledge that creating subtle experiences like suppleness requires expertise and resources (including time) sufficient to create rich, interactive prototypes and iteration. Again, this assumption and the conditions for it are often lacking in educational as well as research contexts. We should revisit prototyping practices—how we teach them and what sorts of things are working now in research and commercial practice.

Different kinds of dialogs with users. Designers hoping to create rich end use qualities like suppleness need to use an expanded repertoire of evaluation techniques—testing in contextually appropriate situations, getting nonverbal, nuanced feedback, probing users in new ways. This is doubly the case if the application area includes social interaction, as the traditional tools for assessing a person’s reaction to an interface often fall short of addressing group response. We need to bring together those who have experimented with new methods for engaging users, with those who are aiming for new use qualities, to produce practices to meet these needs. And we need to shift our view of ‘users’ in the process to co-actors in what we are doing. We are not the first to say this, but nonetheless, it is important. Such a perspective includes thinking of design beyond release to upgrades and a more fluid relationship with user communities.

Drawing upon (and informing) theory in new ways. The designer may need to lean on theoretical structures to guide both design and evaluation in these new domains, that haven’t traditionally been used in HCI—for example, social scientific theory about nonverbal communication and body movement. Theories that deal with real-time details about interaction may be especially helpful. This area of design may also end up pushing basic theory in social science forward to meet new challenges to theories that arise when one can test them with dynamic, adaptive systems. It may be the case with other domains of theory, as well, (e.g. critical and cultural) that we’ll gain new insight and power by building from case studies in which we find strong and credible patterns for designing and achieving specific use qualities. Perhaps we can create a reflected design repertoire, which links theory and design knowledge and pulls these CHI sub-communities together in interesting ways.

More outreach to successful designers about process. CHI researchers should reach out to practitioners who have succeeded and ask them to articulate and share their process, to help us all better understand how to work toward the use qualities we value, and build new collaborations.

In the sharing of results:
We also believe there are changes that could be made in how we share our results, that can help create a stronger dialog and sharing process around qualities like suppleness. We’ve attended events at which practitioners demonstrate their inventions and discuss their work process—e.g. Dorkbot—that provide a strong model for how to include design values such as suppleness in more effective ways in conference presentations. Typically this sort of presentation includes demonstration of the artifact itself working, as well as videos that capture context and prototyping in rich and nuanced ways. The Interaction sessions at CHI are a step toward this, but don’t go far enough, we feel, in challenging/expecting presenters to unveil a fully realized experiential quality, and successfully walk us through the process of achieving that quality (including evaluation with target users). We think it is reasonable to expect those presenting work that has achieved some sort of design value to articulate theoretically-grounded choices they made in prototyping and evaluating, as well.

Ideally, there would be sessions in which one could see accumulated know-how moving into distilled practices and theory for achieving certain use qualities such as suppleness. We could set some desirable use qualities as a community, and ask for submissions that truly achieve these qualities and curate them for re-presentation at the conference. Crucial to creating this type of venue for presentation of work would be a rethinking of what it means to achieve rigor in such work. Reviewers should be encouraged to privilege (and demand) direct experience of the work itself, in order to gauge for themselves experientially if the desired quality has been achieved, and/or dynamic documentation of users that indicates this is the case (this is standard practice in the games community, for example, so it can be done). We believe a re-orientation of reviewer criteria and context for decision-making is crucial here, and is perhaps the single most difficult part of this change to effect.

CONCLUSIONS AND CHALLENGE
This paper has elucidated the use quality ‘suppleness’, as an example of design considerations that emerge as we work to craft interactions that make use of a broader range of human communication capacities in a larger range of contexts, which often include subtle emotional and social qualities. We’ve used suppleness to present some ideas about how HCI design and evaluation practices, as well as practices for sharing our results, may need to change if we truly wish to meet the challenges of these new contexts and tools.

There have been quite a few steps taken already toward putting subtle design qualities further forward in our largest conference, and toward opening up our understanding of what rigor means in evaluating our experiments in these areas. We have advocated in this paper for some additional steps that can be taken. We hope this paper will spark discussion and debate among readers and attendees of CHI 2009, to bring us to a more mature design practice around these qualities, as well as a re-energizing of our research agendas and theoretical constructs.

ACKNOWLEDGMENTS
Sincere thanks to Jarmo Laaksolahti and Jonas Löwgren for valuable feedback on earlier drafts.
REFERENCES


15. Isbister, K., Lawson, S., Ash, J., Nolan, C., and Straus, R. Wriggle! A platform for dynamic and expressive so-