

# On the Naturalness of Touchless: Putting the “Interaction” Back into NUI

KENTON O'HARA, RICHARD HARPER, HELENA MENTIS, ABIGAIL SELLEN, and ALEX TAYLOR, Microsoft Research

Categories and Subject Descriptors: H 5.3 [Information Interfaces and Presentation]: Group and Organization Interfaces – Computer Supported Cooperative Work

General Terms: Human Factors

Additional Key Words and Phrases: Touches interaction, gestural interfaces, natural user interfaces

## ACM Reference Format:

O'Hara, K., Harper, R., Mentis, H., Sellen, A., and Taylor, A. 2013. On the naturalness of touchless: Putting the “interaction” back into NUI. *ACM Trans. Comput.-Hum. Interact.* 20, 1, Article 5 (March 2013), 25 pages. DOI: <http://dx.doi.org/10.1145/2442106.2442111>

## 1. INTRODUCTION

After many decades of research, the ability to interact with technology through touchless gestures and sensed body movements is becoming an everyday reality. The emergence of Microsoft Kinect, among a host of other related technologies, has had a profound effect on the collective imagination, inspiring and creating new interaction paradigms beyond traditional input mechanisms such as mouse and keyboard. Kinect and other technologies form part of the broader suite of innovations that have come to be characterized as Natural User Interfaces (NUI) (e.g., Widgor and Wixon 2011 and Norman [2011]). This moniker includes not only the vision techniques that form that basis of Kinect, but also natural language interfaces, pen-based input, and multitouch gestural input, amongst techniques. The excitement around touchless and body-based interfaces has been accompanied by an increasingly powerful narrative, one that makes the eponymous claim that these new technologies offer an intuitive interface modality, one that does not require users to develop specialist techniques for communicating to computers. What users need to do, instead, is what comes naturally. Consider, for example, the following quote from Saffer.

“The best, most natural designs, then, are those that match the behavior of the system to the gesture humans might actually do to enable that behavior” [Saffer 2009, page 29]

The essential argument is that drawing on existing gestures in everyday life, by identifying the physical movements used to manipulate and understand the world, new interaction paradigms can be developed that will allow people to act and communicate in ways they are naturally predisposed to. They will not have to adapt their action or

---

Authors' addresses: K. O'Hara (corresponding author), R. Harper, H. Mentis, A. Sellen, and A. Taylor, Microsoft Research, Cambridge, UK; email: [oharakenton@gmail.com](mailto:oharakenton@gmail.com).

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies show this notice on the first page or initial screen of a display along with the full citation. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, to redistribute to lists, or to use any component of this work in other works requires prior specific permission and/or a fee. Permissions may be requested from Publications Dept., ACM, Inc., 2 Penn Plaza, Suite 701, New York, NY 10121-0701 USA, fax +1 (212) 869-0481, or [permissions@acm.org](mailto:permissions@acm.org).

© 2013 ACM 1073-0516/2013/03-ART5 \$15.00

DOI: <http://dx.doi.org/10.1145/2442106.2442111>

communications to the peculiarities and limitations of technology; the interface will no longer be a barrier to users, the interface will be them and their gestures.

Such a narrative, of course, does serve a number of purposes: it's good for marketing, for example, making a technology appeal in ways that it might not otherwise do. Many people do not like to use a keyboard, as a case in point, and so Kinect might be especially appealing to them. Such a narrative can also help express high-level visions that set out design and engineering challenges. These can inspire research and development communities not just in HCI, but in hardware and software engineering too; NUI can appeal across the board.

However, elements of this narrative are becoming so deeply embedded in how new forms of interaction are thought about and described that important, albeit apparently minor, distinctions are being elided. Indeed it is not uncommon practice for papers writing about touchless gestural, and body-based interaction to deploy the term *natural* (and its cognate *intuitive*) when characterizing these technologies (e.g., Bhuiyan and Picking [2009], Varona et al. [2008], Corradini [2001], Pavlovic et al. [1997], Baudel and Beaudouin-Lafon [1993], Stern et al. [2008], de la Barré et al. [2005], Wexelblat [1995], Garg et al. [2009], Wu and Huang [1999], Cipolle and Pentland [1998], O'Hagan et al. [2002], Sánchez-Nielsen et al. [2003]). Indeed in a review of 40 years of literature on gesture-based interaction, Karam and Schraefel [2005] cite naturalness as one of the key motivations underlying much of the work in this area. As they say, "much of the research on gesture based interactions claim that gestures can provide a more natural form of interacting with computers." [Karam and Schraefel 2005, page 26]

There are a number of concerns with this treatment to be highlighted here. First of all, gestural interactions are not a homogenous entity. As various authors have articulated, gestural interactions may refer to very different kinds of activities [Quek et al. 2002; Karam and Schraefel 2005]. Based on the work of Quek et al. [2002], Karam and Schraefel [2005] identify different forms of gestural action. These include *deictic* gestures for pointing, *manipulative* gestures that are used to control an object or entity, *semaphoric* gestures that symbolize an object or action with communicative intent, *language* gestures (e.g., sign language), and *gesticulation* or coverbal gestures that accompany speech. These gestural types of course have different properties but at times this is glossed in the literature in the form of conceptual homogenization or where motivations for particular gestural interactions of one type are justified with reference to another type. The notion of *natural* has also been deployed in rather a loose and unquestioning fashion to mean variously, intuitive, easy to use, or easy to learn; these characteristics arising, it is argued through either mimicing aspects of the real world or drawing on our existing tendencies in the areas of communicative, gesticulative, deictic, and manipulative behaviors and actions (see Widgor and Wixon [2011] for a commentary). At times, it is unclear which or all of these characteristics are being alluded to in any particular deployment of the word *natural* and the foundations on which the deployment is made. Aside from this lack of specificity being an important concern itself, many of these basic claims too are being called into question, a notable example here being Norman's [2010] critique of the naturalness of gestural interfaces in terms of their claimed intuitiveness, usability, learnability, and ergonomics.

Norman's critique is indicative of the issue that while using the word *natural* might have become natural, it is coming at a cost. In other words, precisely because the notion of *naturalness* has become so commonplace in the scientific lexicon of HCI, so it is becoming increasingly important, it seems to us, that there is a critical examination of the conceptual work being performed when it is used. There is a need, we contend, to understand the key assumptions implicit within it and how these frame approaches to design and engineering in particular ways. In our view, a close examination of these assumptions will show how they can constrain as much as enable; nuance is required

when thinking about naturalness and this can help refine how touchless gesture- and movement-based applications are used to innovate. In doing this, we want to adopt a somewhat different tack to Norman’s concerns. So while we would agree with Norman’s counter arguments to the various claims of intuitiveness, usability, and learnability that have been applied to gestural interfaces, there is also a sense that such a critique is still operating on the same playing field (albeit on the opposite side) as the proponents of these naturalness claims. That is, attention remains focused on the interface as the potential source of explanation for (or lack of) naturalness, usability, intuitiveness, and learnability. In taking this focus, though, it is our contention that opportunities for better understanding of what can be done with these technologies are sometimes being lost. Broadly speaking, the argument we want to make here is that by situating the locus of naturalness in the gestural interface alone, it is simply being treated as a *representational* concern. But in doing this, attention is perhaps less focused on the in situ and embodied aspects of *interaction* with such technologies. What we want to argue here is that such interactional concerns need to be a more fundamental feature of our discourse and understanding of naturalness and that by doing so, we can better understand the opportunities and constraints for their innovation and adoption.

## 2. REPRESENTATION VS. INTERACTION

The arguments we construct draw from a number of areas. These include the so-called situated interaction literature, going back to the ethnomethodological turn in CSCW (represented in the works of Bannon and Schmidt [1989], e.g., as well as in Suchman [1987], Suchman and Wynn [1984] and many others. For an overview, see Schmidt [2011]). This work largely derives from Garfinkel [1967] and the social, theoretical implications of the later Wittgenstein [1953] (see Button [1991]). This perspective draws attention to the publically available, demonstrative, and “accountable” features of human action. It also draws on phenomenological approaches, represented most famously by Flores et al. (1988) and subsequently in the so-called postphenomenological work of Ihde [2002] and others. This places an emphasis on the body as the source of experiential awareness and subjectivity, and how, through action or praxis, engagement with world comes to be known ([Lave and Wenger 1991]; for a commentary, see Dourish [2001]). The combination of these views can be contrasted with those that tend to be deployed in human factors and ergonomics research which treat the functioning of the brain and the body as specifiable, particularly as this functioning intersects with machinery (see, for instance, Moray [1998]). This view is sometimes called a “positivistic” perspective on action. In similar ways to how these two broad camps have been used to discuss different notions of *context* in ubiquitous computing [Dourish 2004] and notions of *affect* in affective computing [Boehner et al. 2005], we apply the same contrast to thinking about notions of *naturalness* in relations to touchless and body-based interaction.

We begin first with a look at the predominant form of NUI narrative which can, in our view, be considered as grounded in the positivist account of action. In this perspective, the aim of natural interfaces is to leverage and “draw strength from” pre-existing actions that are used in everyday life by people to communicate and to manipulate objects in the world (e.g., Jacob et al. [2008]). The defining idea behind these interfaces, within this perspective, is to make computer interactions through them “more like interacting with the real nondigital world” [Jacob et al. 2008]. Similarly, Abowd and Mynatt argue as follows.

“It is the goal of natural interfaces to support common forms of human expression . . . Humans speak, gesture and use writing utensils to communicate with humans and alter physical artifacts. These natural actions can and should be used as explicit or implicit input to ubicomp systems” [Abowd and Mynatt 2000, pages 30 and 32].

This perspective, then, assumes that existing communicative gestures and actions are pointers toward, and sometimes exact incarnations of, common or even universal “natural interactions”. These interactions are seen as having an ideal, static, and definable state and, though they are not always completely clear or exactly represented in any particular instance, they are something that can be, with sufficient understanding and scientific research, represented and modeled. Such representations and models can ultimately form the basis for defining interfaces to the digital world that will, broadly speaking, mimic their “real-world” counterparts. The naturalness of these interactions is something that is taken as purely a problem of *representation*, ensuring that they are correctly represented in the interaction mechanism itself. In this sense, natural interactions are something detached from the social context in which they might be deployed; they are not constituted by the context, but brought to it.

In characterizing this perspective, our intention is not to critique in a dismissive fashion. Indeed, it is important to acknowledge that such an approach has led to some important successes in terms of interface innovation. Indeed, the suggestion that there are such essential and transsituational phenomena has been a cornerstone of much ergonomics, for example, and this manifests itself in the design of all sorts of contemporary technologies, from kettles to large-scale organizational systems, from cars to airplanes. It also formed the basis of the original HCI work behind the Xerox Star system [Smith et al. 1982]. It is also central to much contemporary analytic philosophy, particularly the philosophy of mind deriving as it does from the causalism avowed by Donald Davidson [1963]. This is showing itself in current manifestations of the theory of embodied cognition, represented in books like Clarke’s *Natural-Born Cyborgs* [2003]. This is also articulated in HCI, though often without the philosophical auspices being made clear (see Hornecker [2005], Hornecker and Buur [2006], and Larssen et al. [2007]). Rather, our intention is to highlight how such a perspective leads interface design in particular directions and this comes at the expense of not taking other directions. Our suggestion is that these other directions (or paths of inquiry) can lead to significant and insightful ways of understanding what human-machine interaction can entail and this innovation can come around touchless gestural and body-based interfaces for computer systems.

As we say this positivist perspective can be contrasted with the situated and phenomenological approaches. Of significance in this general view is a distinction between the *objective* body and the *lived* body (e.g., Merleau-Ponty [1962, 1968]). The objective body can be characterized in terms of how bodily actions might be described from a third person’s point of view; an abstracted description of muscular performance that can be defined and represented. The lived body view, by contrast, concerns the way that people experience and perceive the world through bodily actions. In this perspective, the lived body is in constant rapport with the situated circumstances and it is through actions on the world that these circumstances and the role or function of the embodied actor are made meaningful. The conscious experience of the world and the way it is understood are inseparable from the process of acting in that world. This view emphasizes the subjective construction of meaning through praxis. This subjectivity is, however, publically available; not solely through Husserl’s [1939] technique of introspection, but, as Merleau-Ponty wanted to point out, through everyday practices, such as through discourse.

A second significant element of this perspective comes from Wittgenstein [1967], and his claim that, through action, people create shared meanings with others, and these shared meanings are the essential common ground that enable individual perception to be cohered into socially organized, understood, and coordinated experiences. This draws attention to how actions come to be treated as somehow rational and accountable, as demonstrably about a known-in-common purpose. Garfinkel developed this point and

highlighted how talk, situated talk, or as he put it, reflexive talk, is central to how activities come to be understood. Where Merleau-Ponty emphasized the individual subject and his or her bodily praxis, Wittgenstein (and hence Garfinkel) emphasized the social basis of the individual’s experience, and this pointed towards language and its use in context, to how people act together through talk and other reflexive activities.

Though it would be true to say that there are important distinctions between these two philosophers as there are indeed in the work within HCI that has derived, there is nevertheless a common perspective, particularly when it comes to understanding naturalness and natural gestures or acts. From this view, naturalness is not something to be *represented* but is rather an “occasioned property” of action, something that is actively produced and managed together by people in particular places and at particular occasions, hence the phrase. Of significance here is that these occasioned properties are not just linked to space, to locations of various sorts, but also to the set of persons who occupy these spaces and render them suited for particular actions. Lave and Wenger [1991], along with Brown and Duguid [1991], call these groupings “communities of practice”, by which they mean to highlight how communities cultivate and embody particular sets of skills and know-how, much of which is not articulated through verbal or documented forms but is shared through bodily proximity. Communities make space in this sense or rather make space come to represent and enable embodied learning.

In this respect, the naturalness of how a technology might be interacted with lies not in the physical form of that technology, nor in any predefined interface (natural or otherwise) but in how that form and the interface in question meld with the practices of the community that uses them. This is what is constitutive of “natural use”. It is not technology itself that is natural, but the ways that people can make the actions they perform with technology “apposite”, “appropriate”, or “fitting” to the particular social setting and their particular community. It is in this way that it becomes sensible to say that use is *natural*.

By adopting this perspective, our intention here is explicitly not to use it as a means for justifying why certain types of interactions are more natural than others. Indeed, we would argue that it has been a somewhat unfortunate consequence of how a certain interpretation of the phrase natural interaction has been mobilized in the literature. Instead of being used to help understand how to create and explore more natural-like interfaces, the emphasis on the embodied aspects of action has led some to justify, as a case in point, why tangible computing and body-based interactions “work better” for people because they are “more natural” for people when compared with other forms of interaction. Or, to put this another way, it is sometimes proposed that the success of these systems is because they make better use of users’ kinaesthetic and proprioceptive awareness of their bodies; the systems are thus *more* natural (e.g., Jacob et al. [2008]).

If we engage with the Wittgenstian/Merleau-Ponty perspective, we need to accept that all action is embodied, irrespective of any interaction mechanism or artifact that we may come to use; but we also need to understand that it is through praxis that understanding comes. What is important is both the claim about the centrality of the body as the vehicle for understanding and the *potential for action* [Larssen et al. 2007] that the deployment of the body enables; it is through this that the construction of meaning, sense, and so forth is achieved, that is to say through these actions. This is more than simply a question of material, spatial, and technological determinism whereby our actions are shaped by the material structure of the physical world, then. Rather understanding and meaning of the world are made through the actions being performed.

Articulating these different perspectives is not simply a question of philosophical musing or semantic quibbling. Rather, it serves a very practical purpose of drawing our attention to different ways of understanding touchless and body-based interaction technologies. The positivistic view helps specify what might be designed for; those

Table I. Contrasting Characteristics of Touch vs. Touchless Interaction

Touch	Touchless
co-proximate with surface	distant from surface
transfer of matter	no transfer of matter
pressure on surface	no pressure on surface
momentum of object	no momentum
attrition and wear of surface	no attrition or wear
movement constrained by surface	freedom of movement
haptic feedback	no haptic feedback

gestures constitutive of natural behavior. This view makes investigation of everyday gestures seem like a tractable problem, one that has limits: engineers simply need to build for the vocabulary of known movements. But just as this view makes the engineering seem tractable, so it also tends to close down what might be enabled by natural interaction. It does so because it elides the possibility that what is natural is much more diverse and creatively produced than is suggested by the common use of the phrase natural; different contexts and different communities of practice not only need different forms of NUI, they also sometimes make new forms of “the natural”. In other words, the Wittgenstein/Merleau-Ponty view draws attention to the potential for action enabled by various properties of touchless interaction, and the different communities of practice and settings in which actions are given meaning.

Because so much attention has been given to the positivistic approach to the natural, we turn to discuss how to understand the potential for innovation in this area by looking at the problem from the other view. To do this we shall explore, first of all, the kinds of touchless interactions that one might want to appropriate. We will do this by making a contrast with touch-based systems. We then look at how communities develop and cultivate different needs, and thus come to create contexts for the natural. We then explore how communities and the properties of touchless come to manifest themselves in different real-world contexts, which we illustrate in a series of fieldwork examples.

*Properties of touchlessness.* By starting with properties, it might seem that we are going back on our claim that the naturalness or otherwise of technology is to be understood by reference to a technologies use, rather than being intrinsic to the technology. The properties we want to start with, however, are rather more prosaic features that can be brought to bear in different contexts; nevertheless one can characterize their properties without recourse to context. To help articulate them, we set them out as a series of contrast points with the properties of touch-based interaction (see Table I). This list is not intended to be exhaustive but rather is more indicative of the kind of properties we can attend to (refer to de la Barre et al. [2009]). There are undoubtedly numerous others but what is important is the subsequent way in which these properties are then considered with respect to the different communities of practice and settings.

Let us consider some of these further. The first point of contrast concerns the *proxemic* consequences of touch-based versus touchless interactions (refer to O'Hara et al. [2011], Mentis et al. [2012]). When we interact by touching a system we are required to be coproximate with the surface we are touching: it has to be accessible and open to touch and it has to be in reach. With touchless interaction, by contrast, we can interact at a range of different proximities from the surface of the system. The exact distance from a surface at which touchless interaction can take place depends on the particular sensing technology in question, ranging from a few centimeters to several meters.

The second property we highlight concerns the *transfer of matter* With touch-based interactions, because of the necessity of contact, there is a transfer of matter from the person touching to the device and, vice versa, from the device to the person touching.

Touchless interaction, by contrast, avoids contact and therefore any transfer of matter to or from the system.

Third, in touching something, there is always a certain amount of *momentum and pressure* applied to the surface being touched. This may cause movement, damage, erosion, and attrition. In touchless interaction, by contrast, there is no application of pressure or momentum to the surface in question and therefore no potential for movement damage and erosion.

The fourth property concerns *constraints on movement*. With touch-based interactions, movement is bound and constrained by the shape and properties of the surface being touched. With touchless interaction technologies, by contrast, movement is free and unconstrained by the technology’s surfaces.

Finally, we consider the property of *haptic feedback*. With touch-based interactions, the contact with the surface can provide a rich source of haptic feedback through which manipulations can be finely tuned and refined on a moment-by-moment basis. With touchless interactions, there is an absence of haptic feedback and with that a diminished resource for fine-tuning and refining manipulations in the moment.

For the purposes of simplicity in our argument, we have specified these properties at a particularly high level. For each of these properties it is possible to articulate them at much finer levels of granularity (refer to Rogers and Muller [2006]). Ultimately, the exact level of detail at which we articulate these is done with reference to the potential for action for which we are orienting and the significance of this to certain communities of practice in particular settings.

*Communities of practice.* We turn now to consider Wenger’s [1998] notion of *communities of practice*. What is significant in Wenger’s notion of practice is the coming together of meaning and action. The practices of a particular community are the ways that they experience the world through action and how it is made meaningful. Different properties of an artifact and the potential for action they entail, are seen, interpreted, and made meaningful in different ways by different communities through the ways that they are enacted in their practices. Let us consider, for example, the issue of transfer of matter that takes place due to the contact necessity of touch-based interactions but not for touchless interactions. As Mary Douglas [1966] has eloquently argued albeit *avant le letter* of the term “communities of practice”, people’s orientation towards matter as “clean” or “dirty” is not an inherent, fixed, or absolute classification, but only makes sense with reference to a particular community of practice and the activities in question. Take, for example, scientists and engineers in “clean room” environments and their need to orient to dust particles and other matter in very different ways from other groups. For these scientists, the presence of even the tiniest particle of matter can be sufficient to interfere with carefully planned experiments and manufacturing processes. The meaning of contaminating matter then is very different to this community to what might be considered a contaminant in more every day behaviors and practices. This meaning in turn affects the ways that this community of practice orient towards notions of touch and touchless in the organization of their action. Much of the practices are organized to avoid direct contact with surfaces in ways that risk the transfer of matter. Indeed, the organization of these actions in this way is entirely natural for this group given the particular significance of contamination for them. In this respect, the noncontact property of touchless interaction has a very different meaning-making potential for this community than it does for others. Through this property, an evolving set of practices would be enabled for this community that enable them to experience, interpret, and engage with their world in new ways.

Similar arguments can be applied to the other properties mentioned. Let us consider the issues of pressure and momentum that arises through touch-based interaction but

not present in touchless interaction. Again, if we consider scientists and engineers working in clean-room environments, we can see some very particular ways that this community would orient to such concerns. Scientists and engineers in these clean-room environments, who are working at nano scale, need to orient to movement and vibration in very particular ways. Even tiny vibrations might disrupt experiments and manufacturing processes for these operators that other groups simply would not be concerned with. As before, the particular meaning of pressure, momentum, and vibration for these groups affects the way that activities are organized in relation to touch and not touching. This kind of range in forms of concern—almost a kind of relativity—can be seen in more everyday group concerns with respect to movement and pressure sensitivity of touch. A good example here can be seen in the use of multitouch phones. When held in the hand, the pressure property of touch is not really a worry. But when the same phone is placed on a speaker dock system, the same pressure of touch necessary to control the device puts pressure on the docking socket that, with sustained use, can result in damage both to the phone and docking device. Accordingly, actions are adjusted in such circumstances to avoid potential damage.

*Settings.* Third, the interactional perspective on the naturalness of touchless interaction draws our attention to the settings in which particular communities and groups perform their activities. These settings consist, in part of the physical environment, the architectural arrangement and whole ecology of artifacts within which a piece of interactive technology might be situated. They consist also of a set of other social actors. This physical and spatial structure of these environments then, both enable and constrain how action and practices are organized with respect to information artifacts and other people in the system (e.g., Kendon [2010], Hornecker [2005], O'Hara et al. [2010], Hall [1966], Marshall et al. [2011], Bardram and Bossen [2005]). The features of these settings then can be related to particular properties of touchless interaction.

For example, let us consider the notion of *interaction proxemics* [O'Hara et al. 2010]. This concept labels the spatial consequences of particular interaction mechanisms. For touch-based technologies, the spatial need to be coproximate with the system has consequences for how action can be organized and the particular ways this information can be incorporated into the broader practices within these settings. With touchless interaction, the requirement for proximity to information displays is not there. This different spatial relationship with the information has consequences for when, where, and how this information can be incorporated into the practices in these settings. It changes the relationship between actors and the information and creates different potential for action and meaning making through these interactions. Similarly, if we consider the property of freedom of movement of touchless interaction, it is clear that particular settings may facilitate or enable certain types of gesture and body movement. That is, freedom of movement can be physically hindered by the dimensions of a space, presence of other artifacts, the presence of other people, or the need to concurrently interact with other tools. Again, this affects the potential for how we meaningfully configure action in relationship with these settings.

Also of significance in these settings are particular collaboration and coordination concerns and how the configuration of these activities is achieved in the context of particular interaction possibilities. One might ask how actions are made visible, accountable, and meaningful to the other actors in these settings and how the properties of touchless interaction can be brought to bear in meaningful ways. It will be important to recognize that this will not be a one-way relationship. It is not simply a question of how certain types of interactive gesture or body movements are visible or not to other actors in these settings but also how other features of the coordination and collaborative activities relate to the potential for touchless interaction.

For example, if we consider the need to work in close physical proximity to others in these settings, this may impact on the technical capabilities of the system to track the movements of an individual actor. Different settings too will have particular norms and expectations of appropriate behavior that can be enacted here. The need to attend to these norms and expectations within the social context of these settings imposes important boundaries and constraints on how particular communities orient to specific properties of touchlessness in terms of the movements and actions they perform. It may be entirely appropriate to jump around and wave one’s arms in the comfort of one’s own home but such behavior may be less appropriate for other settings such as the workplace.

### 2.1. Naturalness in Situ

Taking these things together then, what emerges is a different perspective on how we conceive the notion of naturalness in relation to touchless interaction. Naturalness in this perspective is not something that is bound up in a representation of our gesture and body movement; it is not about the ability to infer intent through these representations. It is not simply the exchange of information between man and machine in order to elicit some form of system response. It is not something that can be bound up and packaged solely within the interaction mechanism itself. What is significant about the embodied interaction perspective is how touchless technologies are able to reconfigure our relationship with the material and social world. Naturalness of interactions, in this sense, arises from the potential for action enabled by various properties of touchless interaction and how these properties come to be made meaningful in the practices of specific communities in particular social settings. In designing natural touchless interactions then, our concerns cannot simply be with ever-more enhanced representation and modeling of gesture, movement, and domain physics. These systems should not be judged in terms how well they approximate or fall short of the characteristics of human-human communication. Rather, we need to approach the design of these systems in terms of how they might allow a beneficial reconfiguration of practices and how we experience the world in new ways accordingly.

In order to illustrate these points in a more concrete fashion, we present some fieldwork examples for which we are designing or have deployed touchless interaction technology. The chosen settings are very different in nature, affording us the opportunity to highlight and contrast the occasioned “naturalness” of touchless interaction. It takes different forms in other words, depending upon context.

The first example concerns practices around medical images in surgical settings and opportunities for touchless interaction (e.g., Johnson et al. [2011], Mentis et al. [2012], Wachs et al. [2006, 2007], Stern et al. [2008], Graetzel et al. [2004]). In the second example, we consider practices around an interactive game on a large public screen display (e.g., O’Hara et al. [2008], O’Shea [2009, 2010]).

## 3. TOUCHLESS INTERACTION IN SURGICAL SETTINGS

Our discussion here draws on fieldwork conducted in operating theaters in two large hospitals in the U.K. The observations we undertook covered a variety of different procedures in interventional radiology, neurosurgery, and vascular surgery. Within the theaters there is a wide range of medical imaging equipment and displays used. These allow access to preoperatively captured images such as CT scans and MRI scans, as well as images captured during the course of the procedures such as real-time fluoroscopy and angiographic image sequences. These images are used variously for reference, diagnosis, planning interventions, and for real-time navigation and guidance of equipment on the otherwise hidden inside of the body. The ways that the images need to be viewed, interacted with, and even manipulated is contingent on the particulars of the

procedures in question. Currently within these hospitals, the interactions with these images are achieved through traditional touch-based interaction techniques, primarily keyboard and mouse, but also some use of touchscreens. The purposes of the field-work is to understand how work practices in these settings are currently organized with respect to touch-based technologies with a view to considering opportunities and implications for touchless interaction technology.

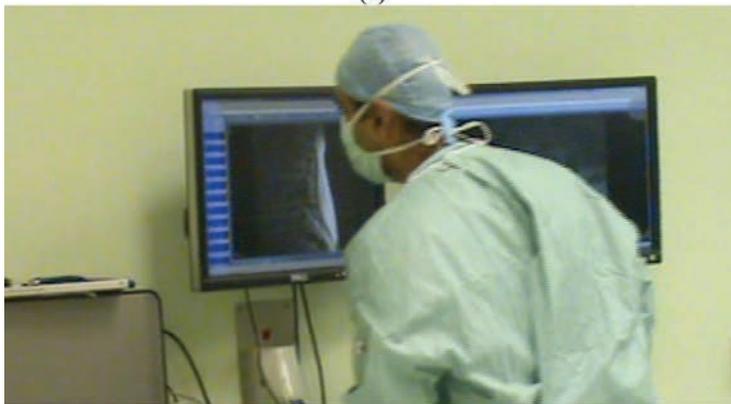
One of the key factors to which people orient in the organization of work in these settings is the boundary between sterile and nonsterile features of the environment. Within these settings, there are areas demarked as sterile and those which are nonsterile. For the members of the surgical team who are scrubbed up (consultant surgeons, radiologists, and scrub nurses) action is organized to avoid contact between sterile and nonsterile surfaces. The interaction technologies used to control the imaging systems in these settings are considered to be nonsterile and therefore not to be touched by the surgeon and others who are scrubbed. Here we see a particular orientation to the transfer of matter that is particular to this group and setting. The transfer of contaminants through touch in this setting means something significantly different to these actors than the ways we might orient to these issues in more everyday circumstances; the notion of what is “dirty” and “clean” is specific to this group. To touch here is a matter of risk to current and future patients as well as staff; it is literally a matter of life and death. This then places restrictions on the surgeon’s interaction with the images. Let us consider an example of how this orientation to the transfer of matter is manifest in practice.

In the scene depicted in Figure 1, we are at the beginning of an open-cut spinal fixation procedure. The consultant surgeon, initially at the patient table, is pressing and massaging the skin around the patient’s spine in order to understand the shape and position of the spinal pedicles. As he is doing this, he is discussing aspects of the spinal curvature with the registrar on the opposite side of the patient table. There remains some uncertainty between them with regard to what they are seeing and feeling on the patient, and as such, they decide to consult the preoperative CT scans of the patient on the PACS display positioned on the wall away from the patient. There is an image already displayed on the screen but this is insufficient to resolve the uncertainty. As such the surgeon needs to select a different view from the CT scans. However, he is unable to touch the nonsterile mouse with the sterile surface of his gloves. He turns and beckons over a nonscrubbed nurse. He says to her: *“Get the mouse and touch the screen there [his finger points at a thumbnail—hovering just centimeters away from the surface of the screen (Figure 1(c))]. . . . that one there, left.”* She asks *“There?”* to which he replies *“Other one.”* While inspecting the images he addresses the registrar and says *“Concave to the right, agreed.”*

Within this sequence, there are a number of important things going on in terms of how the organization is oriented to the issues of touching and not touching. First of all, not touching the mouse is more than simply that: an avoiding the transfer of matter in this particular instance. Not touching is also about demonstrating the ongoing commitment to the unchallengeable delineation between sterile and nonsterile; to making that delineation “real” through “doing”. The second point of significance is how this boundary is managed by an organized distribution of labor between scrubbed (surgeon) and nonscrubbed personnel (nurse) through a process of instruction and pointing. In this particular instance, this kind of activity organization worked effectively. But there are many other times when this way of doing things does not work. For example, on another occasion, the surgeon again had a need to view an image on the PACS display. The system, having being inactive for a period of time, had entered a power-saving mode with the display going to sleep. To reactivate the display and view the image required a simple movement of the mouse. The nurses and operating assistants were all engaged



(a)



(b)



(c)

Fig. 1. A need to reference preoperative images arising in the context of the procedure. The surgeon, being scrubbed up, asks a nurse to manipulate the images on his behalf.

in other activities and so were unable to come and help out. In his frustration, the surgeon lifted his foot up to the shelf in an attempt to jolt the mouse, unsuccessfully. A nurse eventually arrived to help out.

The difficulties with the distribution of labor approach, though, are not simply about uncertain availability of unscrubbed personnel. A further example of image consultation during a procedure reveals more issues. In this example, the procedure involved a particularly difficult spinal fixation surgery on a cancer patient. During the procedure, the surgeon had inserted a rod into the spine but the rod subsequently slipped out of the hole they had drilled into the pedicle. They had to make a decision as to whether to put the rod back in or continue without it. It was apparent to the surgeons that something was just not quite right and, as such, they surgeons returned to the preoperative images. The head surgeon stated the following.

“I was basically looking to see if the anatomy there corresponds to what we found before the X-ray came so I know we are not too high, not too low, and the bone stock, what quality bone we have and that we can do what we set out to do still.”

The responsible surgeon then returned to the table and began to try to insert the wire again. After a little while, he returned to the PACS to view the MRI scans with a surgeon colleague to ascertain if there was a problem with the bone itself. The surgeon points to the PACS and says to the nurse, “Can you just come and reactivate all of this.” The nurse reaches for the mouse and the surgeon begins to point, hovering over icons indicating what to click on. “Over there. Click Ok. Now that.” At this point, there is some uncertainty over the interpretation of the images and whether the bone is diseased or not. One of the surgeons thinks that the images suggest the bone is alright, while the responsible surgeon is “not so sure.” Given the discrepancy in opinion, the surgeon wants to view some alternative images to help resolve the disagreement. He continues to work with the nurse to navigate the images. “Um, no. Go back.” The surgeon realizes the bone density images he needs to see are not available in the PACS system. He turns around to address the rest of the room. “Uh, Who else is here?” He sees the radiographer. “Can you phone your lot downstairs and see if they can recon the bone density scans.” As he turns back to the PACS, the nurse says “Now which one?” He turns his attention back to her. “Um, it is that one?” At this point, the other scrubbed surgeon recognizes the difficulties in communicating the appropriate instructions to the nurse and moves in front of the lead surgeon and says “I’ll do it.” He takes off his glove and thereby sacrifices his sterility<sup>1</sup> in order to be able to navigate through an alternate set of X-ray images as necessary for the medical demands of the situation. The nurse steps back and comments “You are much better to know which one to choose.” The lead surgeon moves beside the second one and looks on as the correct set of images are brought up and flipped through. The two surgeons lean closer to the display and point at a vertebrae as they discuss what they see.”

In part, the problem here stems from difficulties with communicating a more complex set of interaction instructions to the nurse at hand. However, the issue here is more than simply one of communication complexity and lies more with the ways that the surgeons’ professional way of seeing [Goodwin 1994] are inextricably bound up in the active navigation and visualization of the images. With their specific medical knowledge, the “hands-on” progressive stepping through image slices as they view the display *is* their embodied way of seeing, analyzing, interpreting, and decision making. What is revealed

---

<sup>1</sup>It should be noted that the decision to remove the gloves and give up sterility is not lightly taken. The act of rescrubbing is time consuming, taking up to 10 minutes. So giving up this sterility in this way is indicative of the importance of being “hands-on” with the image interactions here.

in this episode is the tension between the need for hands-on control and the need to avoid contamination between sterile and nonsterile surfaces.

Both examples are also of significant in revealing another concern arising from the property of touch-based interaction that requires coproximity to the device. The need for image reference arises while the surgeon is at the patient bedside in the context of what the surgeon sees and feels in the patient and instruments. In order to interact with the images, the surgeons are drawn away from the patient table in order to access the PACS system. They move back and forth between the patient and the images in order to combine what they are seeing on the displays with what they are seeing in the patient. Interpretation of one is done in context of the other. The touch-based nature of the interaction, in part, enforces a separation between these two sources of information and structures the ways they can be combined in the context of collaborative discussion.

*Altering Practice.* What we can postulate here is how touchless interaction, through its potential to work at a distance, could enable the restructuring of these actions. That is, it offers the potential to interact with the images while at the patient table allowing them to be combined in new ways with what the surgeon views and feels at the bedside. This is more than simply doing the same thing but more efficiently; it offers the potential to change the very ways that surgeons are able to perceive and act and the very ways they can perform surgery.

The potential proxemic properties of touchless interaction are made further apparent in the observations of vascular surgery procedures. The settings for these procedures are depicted in Figure 2. As can be seen in the images, the patient table is populated by surgeons, radiologists, and nurses, all of whom are scrubbed. Above the patient table is a bank of displays showing real-time fluoroscopy images and spot images from angiographic runs. The images on these displays are not simply for viewing and navigation but also provide a resource for collaborative analysis and discussion by the team at the patient table. In the context of these discussions, gesturing over the images is an essential feature of the communication that takes place around them. This is significant in itself in relation to touchless interaction. That is, the space in front of the images is already used as a rich site for gesture, albeit gesture for the purposes of communication and collaborative discussion between the surgical team. This is one of the ways they make the images meaningful. The ability for gestural interfaces to become natural in the context of practice in these settings is in part dependent upon how they cohabit the space in front of displays already used for gestures, gestures which gain their meaning from the context.

But there are additional features of gesturing and pointing behaviors in the context of talk around these images that warrant discussion here. In the sequence in Figure 2, the surgeon was discussing the fluoroscopy images with the radiologist. The images were not perfect presentations of the anatomy and sometimes not all the vessels were clearly visible. It became a matter of interpretation as to precisely what bit of the anatomy was being seen. In looking at the images, the surgeon wanted to make sure that he had come out of the origin of celiac artery but was uncertain from the image whether this was the vessel in view. The consultation with the radiologist was to resolve this uncertainty. The significant feature of this episode is the ways that the surgeon and the radiologist attempted to get closer to the image with their pointing. The nature of the query involved a precise delineation of particular features of the image while they spoke. Getting closer to the image with their pointing is what enabled them to more precisely resolve the reference points in the context of their talk. Gesturing from a distance did not enable the precision of pointing. Getting closer to the image in this way is not always easy and may involve awkwardly leaning over the patient. For the surgeon, it was not possible to lean across sufficiently. In order to overcome this,



(a)



(b)



(c)

Fig. 2. Ways of getting closer to the image for deictic reference.

he grabbed a catheter wire from the surgical table which was rigid enough and long enough to act as a pointer to the image. In Figure 2(c), the surgeon gestured with the catheter (circled in red) and said, “*You can see it coming off there can’t you?*”

Getting closer to these images in various different ways enabled them to be made meaningful in the context of collaborative discussion. Again we can reflect on potential opportunities for action here that relate to the proxemic qualities of touchless interaction. Using touchless image tracking to control some form of pointer on the image displays would be an interesting way of exploiting the proxemic qualities of touchless interaction to enable the surgeons to get “closer” to image features. It is this potential that would allow them to create new meanings through touchless engagement and it is in this meaning making that the interactions might come to be realized as natural.

In sum then, what we aim to have highlighted in these examples are some of the ways that surgical teams within these settings organize their behavior with respect to the touch-based interaction mechanisms currently used to interact with images in the operating theater. These episodes aim to show how social action is organized and oriented to some key properties of touch and touchlessness, in particular, the issues of transferring matter through touch and the proxemic consequences of touch. Understanding the ways that behavior is organized around these issues allows us to postulate how touchless interaction might enable beneficial restructuring of action in these settings, for example, by allowing surgeons to have “hands-on” control without the need for contact, or by enabling new spatial relationships with the medical images and patient and hence hands-on control. It is through these potential transformations of practice and production of meaning that the naturalness of touchless interaction will be found in these settings for the surgical team in these settings.

#### 4. COLLABORATIVE PLAY IN AN URBAN SCREEN GAMES

We now turn to a second example that highlights some of the different ways in which touchlessness is adopted in the production of social order and meaning. In this example, we consider embodied interaction with large public display applications in a city environment. We base our discussion around three interactive games deployed on the BBC’s network of large public screens (approximately 5m × 5m) installed in cities across the U.K.

The games in question are depicted in Figure 3. The first game is called the *Red Nose game* (see O’Hara et al. [2008] for further details). In this game, a group of red clown noses appear on screen. When one nose touches another nose, they merge together to form a larger nose. The aim is for players to move all the noses together until they have all merged into one large single nose. A camera pointing away from the screen toward the players in front of the screen captures a moving image of the players. The image-processing algorithm for the game performs simple edge detection on objects in the camera view. When the edge of a person or object contacts the outline of a red nose, simple physics are adopted such that the nose moves in the direction that the contacting edge “pushes”. In this respect, player movements in front of the screen move the noses *touchlessly*. Importantly, there are no predefined movements or gestures that the system recognizes and interprets. Rather, the players can determine exactly how to collaboratively organize their movements to achieve a particular social effect and within the particular social constraints of the setting.

The second game in question is *Hand from Above* [O’Shea 2009]. The installation again uses the camera above the screen pointing out toward the area in front of the screen. In the game, a large hand appears on the screen that moves toward the detected edges of people in the camera image. The hand then performs one of a variety of actions such as tickling, squashing, or flicking the image of the detected person across the screen.



(a)



(b)

Fig. 3. Three interactive games played on BBC public displays. The games are (a) Red Nose game; (b) Hand from Above (reproduced with kind permission from Chris O'Shea - [www.chrisoshea.org/hand-from-above](http://www.chrisoshea.org/hand-from-above)); (c) Question of Sport Relief.



(c)

Fig. 3. (Continued)

The third game is *Question of Sport Relief*, a multiple choice quiz game based around the television quiz show *Question of Sport* [O’Shea 2010]. In this game, a question is presented to the big screen audience in front of the display. Four answers are then presented, one in each of the four quadrants of the display. To choose an answer, the players move so that their image appears in the quadrant corresponding to that answer. Once in position, then need to move around in that space. The more movement that is registered by the software, the more the particular “power bar” associated with that answer increases. The answer with the highest power at the end of a countdown is the one that is selected.

In thinking about touchless interactions with these large displays there is the obvious pragmatic concern around the constraints of physical reach. That is, with the scale of these displays, it is simply not possible to physically reach all parts of the screen in order to enable interaction with the on-screen objects. Therefore, the proxemic qualities of touchless interaction provide a means by which such constraints can be overcome. But of greater significance with all these games is the public context in which such interactions are taking place. People may come to these places as individuals or with family or friends. But they are also there with the larger community of people in the vicinity with whom they are unacquainted. It is in enacting out these various relationships that the interactions with the system are made meaningful. This has particular implications for the ways that interactions with the system come to be organized and made natural. Let us consider some examples.

In the Red Nose game the public nature of the interactions at times created a certain evaluation apprehension that inhibited participation by certain people. There was a particular reluctance, for example, to be the first or only person playing the game while being watched by others. People would join in if other people were already playing, or would play if other members of their immediate group accompanied them or egged them on from the sidelines. Of particular interest here was that adults would join in with

their children but would be less likely to play on their own. Here enacting the parent-child relationship became a means by which their movement-based interactions were made accountable and justifiable. It helped make what might be considered slightly curious behavior in these settings, understandable to the watching public. These kinds of social concerns can also be seen in people's interactions with Hand from Above game. For example, people would sometimes run away from the hand in order to avoid being the public spectacle. In another instance where two young friends were in front of the game, one of the friends pushed the other in an attempt to get her tickled by the hand. The girl resisted and quickly withdrew back to the safety of being in close proximity to her friend. What is significant here is how these actions are organized around specific features of the system and setting to enact a particular relationship. In the first instance, the actions were designed to playfully isolate and embarrass the one girl and in the second instance for her to withdraw back to the safety of being with a friend.

This apprehension in public was also apparent in the Question of Sport Relief game. In particular here was the tendency to follow the crowd and assemble on a single answer. Even when the answer was wrong, people would join with the crowd rather than stand isolated on a different answer. Here, then, we see social concerns shaping the nature of these interactions. The close clustering in a crowd in this game is of significance too with regard to the movements performed. Individuals in this context felt more anonymized. Acting as a unit in this way can remove some of the inhibitions of individual public performance in ways that socially facilitates movement.

The effect of these public settings on the organization of these interactions, though, is not simply one of social inhibition. By contrast, the public nature of these interactions was also employed as an opportunity to perform and show off. Of significance here is how the different ways of implementing touchless interaction in these games enabled a certain *expressive latitude* in the organization of movement [Bowers and Hellstrom 2000; Larssen et al. 2004]. We see a good example of this behavior in the Red Nose game involving a young teenage male player. The boy was there with a group of friends who were watching him while he played the game. In playing the game, the boy used very exaggerated and acrobatic movements in order to move the noses and would frequently turn round toward his friends in order to get their acknowledgement. They would cheer him on as he performed more elaborate moves. What we see here is how the movements are designed not simply to interact with the system. Rather they are also designed to be a performance to his watching friends. The friends, in cheering, also further encouraged these kinds of movements. It is in this relationship that they are made natural and meaningful. Importantly in the Red Nose game, the implementation is not about inferring communicative intent on the part of the player. There are no predefined set of gestures to be "interpreted" by the system. Rather, the simple physics in the system provides a freedom within which the movements and interaction can be shaped in the enactment of particular social relationships.

We can see examples of this too in the Hand from Above game. Here the relationship with the system is more curious in that the movements are performed not so much as a means of controlling the system but rather are designed in response to the way the system behaves. A nice illustration here is how people would position themselves, bend over, and wiggle their bottom for it to be "tickled" by the giant hand on the screen. These kinds of behaviors were performed with humorous intent to make their friends and other spectators laugh. Similarly, some people would run around in attempt to be chased by the hand, again as a performance for others around. The naturalness of these interactions thus was not so much bound up in the system, but in the ways that they were mobilized for particular social effect in these settings.

The Question of Sport Relief game also offers some different possibilities for the performative aspects of relationship work. A key feature of the way touchless interaction was implemented in this game concerned its response to movement. Essentially the system was designed to respond simply to any changes in the image. Any form of movement could effect such a change in the image. The more the image changes over time, the better the response. As such, there is again no predefined or interpreted set of gestures or movement encoded in the system. People would jump up and down as a crowd, frantically waving their arms. At times, people would also put their arms around each other and chant as they jumped up and down in unison. Putting arms around others in this context was not a question of interaction with the system. Rather, it was a particular form of relationship work played out through the interactions with the system. The crowd behaviors were given meaning through these interactions with the system; putting arms around strangers in this context was meaningful in turn because of the meaning provided by the presence of the system through the interactions; being in such close proximity to strangers was given meaning through these interactions. In another interesting example here, a father would repeatedly throw his baby into the air and catch him in order to register movement with the system. The father here was doing more than just interacting with the system. He was using the properties of the touchless implementation to create fun movement for the child. It is in the context of this parent-child relationship that the movements are given meaning.

## 5. DISCUSSION

Through these different fieldwork settings and applications, we can start to build up a picture of the varied ways that touchless technologies might acquire significance in everyday contexts and setting by diverse communities of people. It is with these in mind that we can now revisit some of the initial arguments set out in the earlier part of the article regarding the naturalness of these kinds of technologies and what this means for how we might approach their design. In setting out this discussion, our aim is not to provide a prescriptive set of rules and design guidelines for gesture-based interaction and indeed, in light of the preceding examples, it is not clear that this would be an entirely tractable undertaking. Rather, it is to reorient the designers of such systems to an important set of additional concerns than are not readily apparent in the current conceptualizations of *natural* interaction. These current conceptions, we have argued, have adopted a broadly positivist viewpoint of natural interaction. Within this positivist viewpoint, natural interaction has come to mean various things such as intuitive, easy to use, and easy to learn. While of course these can be regarded as important characteristics of any good design, there is an additional layer of narrative that is often present regarding the source of these characteristics. This is manifest in references to things such as the human tendency to communicate with various forms of semaphoric, gesticulative, and deictic gestures or may include reference to the ways that we use gesture and actions to physically manipulate objects in the world. In this respect, there are considered to be natural ways of being; states and actions that can be defined and represented. By leveraging these preexisting forms of action and communication in our approach to interaction design, the argument is that we can make them more intuitive, easy to use, more learnable, that is, more natural forms of interaction. Naturalness in this respect is treated as a purely representational concern that is bound up within the interface itself.

These particular aspects of “naturalness” have already been directly called into question by the likes of Norman [2010]. Similarly, while there are often some reasonable arguments made for reality-based interfaces (e.g., Jacob et al. [2008]), there are equally

good arguments to the contrary that express some of the potential limitations of these approaches, for example, where the possibilities of digital interaction extend beyond any meaningful counterpart in the real world (e.g., Hollan and Stornetta [1992]).

Our intention in this article, however, has been to offer a different form of critique of the representational account of naturalness by drawing on the theory of embodied interaction and the phenomenological, Wittgenstinian, and situated action ideas from which it is derived. From this perspective, the representational accounts of naturalness can be seen to focus primarily on the *objective body*, whereby our movements and gestures can be characterized simply in terms of their muscular descriptions, selecting and then representing the right set of muscular descriptions will lead to naturalness. What is missing, though, in this focus on the objective body is any reference to the *lived body* which concerns how we experience the world through our gestures and actions and how the bodily actions of the embodied actor are made subjectively and socially meaningful. The objective body then is only a partial account of human action, which has led to an overly narrow set of concerns in the way we understand and characterize gesture and action. This in turn leads to a narrow research agenda in which ever-richer representations of action and naturalness in terms of intuitiveness, ease of use, and learnability become ends in themselves. These are not unimportant but this focus underplays a whole host of concerns that become revealed with stronger reference to the socially lived body of the embodied actor.

These concerns become apparent when we consider these issues in the context of the fieldwork examples. For example, in the surgical episodes, the chief motivations driving the development of touchless gestural interactions with imaging equipment in these areas are not about the development of more intuitive and easy-to-use interfaces. Indeed, the surgeons have developed considerable expertise with these systems and are very adept at using the more traditional mouse and keyboard to interact with the medical images in particular ways. As such, it is difficult to argue for the shift to gestural interaction purely on the grounds of any claimed benefits of naturalness and intuitiveness associated with them. It also doesn't make any clear sense to base any new forms of interaction around the current gestures and actions they perform. What is driving the development of systems such as those deployed at Sunnybrook Hospital in Toronto<sup>2</sup> and our own system development<sup>3</sup> is a rather different set of concerns associated with the issue of sterility. The difficulties with current touch-based interaction techniques lie in the particular constraints they impose on interaction by a scrubbed surgeon in these settings. Within these settings, certain objects come to be designated as sterile or nonsterile, which in turn affects the organization of action by scrubbed and nonscrubbed personnel within the operating theater. Scrubbed surgeons are unable to touch the imaging systems in the nonsterile areas, which affects the ways that they are able to mobilize such resources in the context of their surgery. For example, they may have to move away from the patient and subsequently undertake a time-consuming rescrubbing process after any interaction. Or they may have to instruct an unscrubbed assistant to conduct the manipulations on their behalf. Not only is this cumbersome but removes any direct control by the surgeon and thereby interferes with both their professional vision and the ways the imaging resources are mobilized in the context of collaborative discussion.

What is of significance about touchless gestural-based systems for these people in these contexts lies in the new possibilities for the *lived body*. By adopting touchless gestures, the surgeon is able to regain control over the manipulation of images that is so

---

<sup>2</sup><http://sunnybrook.ca/uploads/N110314.pdf>

<sup>3</sup><http://research.microsoft.com/en-us/projects/touchlessinteractionmedical/>

vital to his or her ability to interpret and analyze these images; it is fundamental to his or her professional vision. The touchless gestures also provide his or her surgeons with the ability to manipulate and gesticulate over the images from a distance. This then entails new possibilities for how the surgeons are able to spatially organize themselves with respect to the imaging systems, with respect to the other team members, and with respect to the patient. They are no longer forced to move away from the patient bedside when dealing with images allowing them to be mobilized in the context of what they are seeing and doing to the patient's body. They no longer need to stretch over a patient in order to operationalize deictic actions around the images. The significance of these systems, then, lies in how new potentials for action allow the work practices to be reconfigured in ways that are meaningful to these particular communities and settings. Through these new practices and interactions, such systems come to be rendered natural. It is adopting the perspective of embodied interaction and its focus on the lived body that these concerns can be made visible. And it is these types of concern that arguably should be more central to our design and innovation agenda in this area as opposed to a simplistic adherence to reality-based naturalness and intuitiveness of interfaces.

Similar arguments can be made with respect to the urban screen examples, albeit with a different set of meanings and issues being highlighted. Again, if we adopt the positivist representational approach to the naturalness of these interactions, all we end up with is a rather crude ability to understand what is going on with these applications in the settings and consequently a rather sparse resource for thinking about their design or why particular approaches to representation work well. So, for example, with the Red Nose game, the adoption of a simple edge detection tracking mechanism has particular consequences that extend beyond notions of intuitiveness. So while such a representational approach renders the control of the red noses reasonably intuitive, easy to learn, and use, this in itself is not all that interesting. Arguably this immediacy of the system has particular significance in these public settings where opportunities for interaction are fleeting and where investment in developing expertise is less of an option than, say, in the living room. However, where the real significance of these representational choices is situated is in the enactment of particular social relationships in these settings, whether these be with known others or strangers. It is the flexibility with which this particular representational choice enables the nuanced and enactment of relationships in these settings that is key here. A single person is treated as amorphous blob in much the same way as a closely huddled group of people is treated as an amorphous blob with an edge around it. For a collection of strangers playing the game, we can see how people can spatially configure themselves at a socially safe distance from each other as appropriate for the relationship at hand. At the same time we can see how this enables familiar people to huddle close together and play for a variety of situated purposes such as to achieve social closeness, to overcome social inhibitions of performing alone in public, or to render one's actions accountable to those around, such when an adult plays the game with a child. The simplicity of this representational approach is also of significance in highlighting the performative meaning of actions and gestures in these settings, enabling a certain expressive latitude in the gestures and actions that was again occasioned. Action and gestures were designed both to show off to watching friends or to be less visible for those with greater social inhibitions. Similar issues are at play in the other game examples. In the Question of Sport Relief game, the system simply responds to the frequency of change of pixels from one video frame to the next. But what is again of significance here is not an objective characterization of action but the lived bodily experiences that are enacted through these actions. So people were able enact different actions such

as configuring themselves closely as a crowd, moving energetically together in unison, or doing things such as throw a child into the air in an enactment of the parent-child relationship. The occasioned nature of these actions is of course more rich and nuanced than the high-level characterizations we are presenting here but the point is that this occasioning is key to understanding the significance of these particular forms of interaction.

In sum, then, our aims in this article have been to open up the discussion around the naturalness of touchless and gesture-based interaction by drawing on the theories of embodied interaction and situated action. In this perspective, naturalness is not something that lies purely within the interface itself and is not something that can be treated simply as a representational concern through which intuitiveness ease of use, and learnability can be achieved. From a design perspective, such a representational account of naturalness with respect to these systems is in fact rather pernicious and only serves to focus our attentions on narrow set of concerns with the objective body. By adopting the perspective of embodied interaction, what we hope to have shown is that there are a broader set of concerns, beyond the objective characterization of the body, that relate to the lived bodily experiences of the embodied actor interacting with these systems. Naturalness, here, is an occasioned property of action that social actors actively manage and produce together in situ through their interaction with each other and the material world. Of importance are the ways gestures and actions are performed and made *meaningful* in particular social settings through which naturalness is achieved. Our attention in understanding the naturalness of these interactions is drawn to their particular properties and what these might mean for particular communities of practice in certain settings.

From a design perspective, then, the concerns of our approach are not framed so much as a problem of human-machine communication, that is one of how can the system better understand what we are trying to do or how we can make it easier for us to communicate with the system. Rather, our concerns lie in how the properties of the technology and the social system are combined together in the production of meaningful and natural interaction. Importantly this coming together is more than the simple material determinism that is apparent in some of the ways that ideas from embodied interaction have been adopted. That is, we are not arguing that the technology and material world somehow constrain and shape our actions in socially meaningful ways. Rather, it is in our coming together with the technology and material world that our interactions become configured in new and meaningful ways. Touchless interactions are not just ways of carrying out the same things but in a new way. Rather, they change the ways that we perceive and understand the world through the embodied actions that we are able to perform.

## REFERENCES

- ABOWD, G. D. AND DIX, A. J. 1994. Integrating status and event phenomena in formal specifications of interactive systems. In *Proceedings of the 2<sup>nd</sup> ACM SIGSOFT Symposium on Foundations of Software Engineering*. Addison-Wesley/ACM Press.
- BANNON, L. AND SCHMIDT, K. 1989. CSCW: Four characters in search of a context. In *Proceedings of the 1<sup>st</sup> European Conference on Computer Supported Cooperative Work (ECSCW'89)*. 358–372.
- BARDRAM, J. AND BOSSEN, C. 2005. Mobility work: The spatial dimension of collaboration at a hospital. *J. Comput. Supported Coop. Work* 14, 2.
- BAUDE, T. AND BEAUDOUIN-LAFON, M. 1993. Charade: Remote control of objects using free-hand gestures. *Comm. ACM* 36, 7, 28-35.
- BOEHNER, K., DEPAULA, R., DOURISH, P. AND SENEGERS, P. 2005. Affect: From information to interaction. In *Proceedings of the 4th Decennial Conference on Critical Computing: Between Sense and Sensibility (CC'05)*, O. W. Bertelsen, N. O. Bouvin, P. G. Krogh, and M. Kyng, Eds., ACM Press, New York, 59–68.

- BOWERS, J. AND HELLSTRÖM, S. O. 2000. Simple interfaces to complex sound improved music. In *CHI'00 Extended Abstracts on Human Factors in Computing Systems*. ACM Press, 125–126.
- BHUIYAN, M. AND PICKING, R. 2009. Gesture control user interface, what have we done and what's next? In *Proceedings of the 5th Collaborative Research Symposium on Security, E-learning, Internet and Networking (SEIN'09)*.
- BROWN J. S. AND DUGUID, D. 1991. Organizational learning and communities-of-practice: Toward a unified view of working, learning, and innovation. *Organ. Sci.* 2, 1, 40–57.
- BUTTON, G. 1991. *Ethnomethodology and the Human Sciences*. Cambridge University Press.
- CIPOLLE, R. AND PENTLAND, A. 1998. *Computer Vision and Human-Computer Interaction*. Cambridge University Press.
- CLARK, A. 2003. *Natural-Born Cyborgs: Minds, Technologies and the Future Human*. Open University Press, Milton Keynes.
- CORRADINI, A. 2001. Real-Time gesture recognition by means of hybrid recognizers. In *Proceedings of the Revised Papers from the International Gesture Workshop on Gesture and Sign Languages in Human-Computer Interaction (GW'01)*. I. Wachsmuth and T. Sowa, Eds., Springer, 34–46.
- DAVIDSON, D. 1963. Actions, reasons, and causes. In *Essays on Davidson: Action and Events*, Oxford University Press, 1980, 3–20.
- DE LA BARRE, R., CHOJECKI, P., LEINER, U., MUHLBACJ, L., AND RUSCHIN, D. 2009. Touchless interaction—novel cases and challenges. In *Proceedings of the International Conference on Human Computer Interaction (HCI'09)*, J. Jacko, Ed., Springer, 169–169.
- DE LA BARRÉ, R., PASTOOR, S., CONOMIS, C., PRZEWOZNY, D., RENAULT, S., STACHEL, O., DUCKSTEIN, B., AND SCHENKE, K. 2005. Natural interaction in a desktop mixed reality environment. In *Proceedings of the 6<sup>th</sup> International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS'05)*.
- DOURISH, P. 2001. *Where the Action Is: The Foundations of Embodied Interaction*. MIT Press.
- DOURISH, P. 2004. What we talk about when we talk about context. *Pers. Ubiquitous Comput.* 8, 1.
- DOUGLAS, M. 1966. *Purity and Danger*. Routledge, London.
- FLORES, F., GRAVES, M., HARTFIELD, B., AND WINOGRAD, T. 1988. Computer systems and the design of organizational interaction. *ACM Trans. Inf. Syst.* 6, 2.
- GARFINKEL, H. 1967. *Studies in Ethnomethodology*. Prentice Hall.
- GARG, P. AGGARWAL, N., AND SOFAT, S. 2009. Vision based hand gesture recognition. *World Academy Sci. Engin. Technol.* 49.
- GOODWIN, C. 1994. Professional vision. *Amer. Anthropol.* 96, 3, 606–633.
- GRAETZEL, C., FONG, T., GRANGE, S., AND BAUR, C. 2004. A non-contact mouse for surgeon-computer interaction. *Technol. Health Care* 12. IOS Press.
- HALL, E. T. 1966. *The Hidden Dimension*. Doubleday, New York.
- HOLLAN, J. AND STORNETTA, S. 1992. Beyond being there. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI'92)*. 119–125.
- HORNECKER, E. AND BUUR, J. 2006. Getting a grip on tangible interaction: A framework on physical space and social interaction. In *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems (CHI'06)*. ACM Press, New York, 437–446.
- HORNECKER, E. 2005. A design theme for tangible interaction: Embodied facilitation. In *Proceedings of the 9<sup>th</sup> European Conference on Computer Supported Cooperative Work (ECSCW'05)*.
- HUSSERL, E. 1939. *Experience in Judgement*. Routledge, London.
- IHDE D. 2002. *Bodies in Technology*. University of Minnesota Press, Minneapolis, MN.
- JACOB, R., GIROUARD, A., HIRSHFIELD, L., HORN, M., SHAER, O., SOLOVEY, E., AND ZIGELBAUM, J. 2008. Reality based interaction: A framework for post-WIMP interfaces. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI'08)*.
- JOHNSON, R., O'HARA, K., SELLEN, A., COUSINS, C., AND CRIMINISI, A. 2011. Exploring the potential for touchless interaction in image-guided interventional radiology. In *Proceedings of the Conference on Human Factors in Computing*. 3323–3332.
- KARAM, M. AND SCHRAEFEL, M. 2005. A taxonomy of gestures in human-computer interaction. Tech. rep. ECSTR-IAM05-009, Electronics and Computer Science, University of Southampton.
- KENDON, A. 2010. Spacing and orientation in co-present interaction. In *Proceedings of the COST Training School*. Springer, 1–15.
- LAVE, J. AND WENGER, E. 1991. *Situated Learning: Legitimate Peripheral Participation*. Cambridge University Press.

- LARSSSEN, T., ROBERTSON, T. AND J EDWARDS, J. 2007. The feel dimension of technology interaction: Exploring tangibles through movement and touch. In *Proceedings of the Conference on Tangible and Embedded Interaction*. 271–278.
- LARSSSEN, A. T., LOKE, L., ROBERTSON, T. AND EDWARDS, J. 2004. Understanding movement as input for interaction –A study of two Eyetoy (TM) Games. In *Proceedings of the Annual Meeting of the Australian Special Interest Group for Computer Human Interaction (OZCHI'04)*.
- MARSHALL, P., ROGERS, Y., AND PANDITI, N. 2011. Using f-formations to analyse spatial patterns of interaction in physical environments. In *Proceedings of the Conference on Computer Supported Cooperative Work (CSCW'11)*. 445–454.
- MENTIS, H., O'HARA, K., SELLEN, A., AND TRIVEDI, R. 2012. Interaction proxemics and image use in neurosurgery. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*.
- MERLEAU-PONTY, M. 1962. *Phenomenology of Perception*. Routledge, UK.
- MERLEAU-PONTY, M. 1968. The intertwining - the chiasm. In *The Visible and the Invisible*. Northwestern University Press, Chicago, IL.
- MORAY, N. 1998. Identifying mental models of complex human-machine systems. *Int. J. Ind. Ergonom.* 22, 4–5, 293–297.
- NORMAN, D. 2010. Natural user interfaces are not natural. *Interact.* 17, 3, 6–10.
- O'HAGAN, R., ZELINSKY, A., AND ROUGEAX, S. 2002. Visual gesture interfaces for virtual environments. *Interact. Comput.* 14, 231–250.
- O'HARA, K., KJELDSKOV, J., AND PAAY, J. 2011. Blended interaction spaces. *ACM Trans. Comput.-Hum. Interact.* 18, 1.
- O'HARA, K., GLANCEY, M., AND ROBERTSHAW, S. 2008. Collective play in an urban screen game. In *Proceedings of the Conference on Computer Supported Cooperative Work (CSCW'08)*.
- O'SHEA, C. 2009. Hand from above. <http://www.chrisoshea.org/hand-from-above>
- O'SHEA, C. 2010. Question of sport relief. <http://www.chrisoshea.org/big-screen-quiz>
- PAVLOVIC, V., SHARMA, R., AND HUANG, T. 1997. Visual interpretation of hand gestures for human-computer interaction: A review. *IEEE Trans. Pattern Anal. Mach. Intell.* 19, 7, 677–695.
- QUEK, F., MCNEILL, D., BRILL, R., DUNCAN, S., MA, X.-F., KIRBAS, C., MCCULLOUGH, K. E., AND ANSARI, R. 2002. Multimodal human discourse: Gesture and speech. *ACM Trans. Comput.-Hum. Interact.* 9, 3, 171–193.
- ROGERS, Y. AND MULLER, H. 2006. A framework for designing sensor-based interactions to promote exploration and reflection in play source. *Int. J. Hum.-Comput. Stud.* 64, 1.
- SAFFER, D. 2009. *Designing Gestural Interfaces*. O'Reilly.
- SÁNCHEZ-NIELSEN, E., ANTÓN-CANALÍS, L., AND HERNÁNDEZ-TEJERA, M. 2003. Hand gesture recognition for human-machine interaction. *J. WSCG* 12, 1–3.
- SCHMIDT, K. 2011. *Cooperative Work and Coordinative Practices: Contributions to the Conceptual Foundations of Computer-Supported Cooperative Work*. Springer.
- SMITH, D. C., IRBY, C., KIMBALL, R., AND HARSLEM, E. 1982. The star interface: An overview. In *Proceedings of the National Computer Conference (AFIPS'82)*. R. Brown, and H. Morgan, Eds., 515–528.
- STERN, H., WACHS, J., AND EDAN, Y. 2008. Optimal consensus intuitive hand gesture vocabulary design. In *Proceedings of the IEEE International Conference on Semantic Computing (ICSC'08)*. IEEE Computer Society, 96–103.
- SUCHMAN, L. 1987. *Plans and Situated Actions: The Problem of Human-Machine Communication*. Cambridge University Press.
- SUCHMAN L. AND WYNNE, E. 1984. Procedures and problems in the office. *Office Technol. People* 2, 133–154.
- VARONA, J., JAUME-I-CAPÓ, A., GONZÁLEZ, J., AND PERALES, F. J. 2008. Toward natural interaction through visual recognition of body gestures in real-time. *Interact. Comput.* 21, 1, 3–10.
- WACHS, J., STERN, H., EDAN, Y., GILLAM, M., FEIED, C., SMITH, M., AND HANDLER, J. 2006. A real-time hand gesture interface for medical visualization applications. <http://web.ics.purdue.edu/~jpwachs/papers/WSC10.wachs.pdf>
- WACHS, J., STERN, H., EDAN, Y., GILLAM, M., FEIED, C., SMITH, M., AND HANDLER, J. 2007. Real-Time hand gesture interface for browsing medical images. *IC Med.* 1, 2.
- WENGER, E. 1998. *Communities of Practice: Learning, Meaning and Identity*. Cambridge University Press.
- WEXELBLAT, A. 1995. An approach to natural gesture in virtual environments. *ACM Trans. Comput.-Hum. Interact.* 2, 3, 179–200.
- WIDGOR, D. AND WIXON, D. 2011. *Brave NUI World: Designing Natural User Interfaces for Touch and Gesture*. Morgan Kaufmann.

WITTGENSTEIN, L. 1953. *Philosophical Investigations*. Blackwell, Oxford, UK.

WITTGENSTEIN, L. 1969. *On Certainty*. Blackwell, Oxford, UK.

WU, Y. AND HUANG, T. 1999. Vision-Based gesture recognition: A review. In *Proceedings of the International Gesture Workshop on Gesture-Based Communication in Human-Computer Interaction (GW'99)*. A. Braffort, R. Gherbi, S. Gibet, J. Richardson, and D. Teil, Eds., Springer, 103–115.

Received October 2011; revised March 2012; accepted July 2012