Do Architects and Designers Think about Interactivity Differently?

DAVID KIRSH, UCSD

This essay has three parts. In Part 1, I review six biases that frame the way architects and human–computer interaction (HCI) practitioners think about their design problems. These arise from differences between working on procedurally complex tasks in peripersonal space like writing or sketching and being immersed in larger physical spaces where we dwell and engage in body-sized activity like sitting, chatting, and moving about. In Part 2, I explore three types of interface: classical HCI, network interfaces such as context-aware systems, and socio-ecological interfaces. An interface for an architect is a niche that includes the very people who interact with it. In HCI, people are still distinct from the interface. Because of this difference, architectural conceptions may be a fertile playground for HCI. The same holds for interactivity. In Part 3, I discuss why interactivity in HCI is symmetric and transitive. Only in ecological and social interaction is it also reflexive. In ecological interfaces, people co-create bubbles of joint awareness where they share highly situated values, experience, and knowledge.

1 INTRODUCTION

University curricula in Architecture and human–computer interaction (HCI) are different, so much so that it is hard to find evidence of cross-pollination. Why is that? Is it because one field is about designing buildings where people dwell, while the other is about making interfaces for people to control things or to receive services?

In what follows, I argue that the conceptual separation between the two fields is more interesting than the historical distinction between inhabiting vs. controlling, though admittedly it is related. I distinguish six ways architects and HCI practitioners think differently about their design problems, and that, moreover, these six are themselves a reflection of a fairly profound division in how the two disciplines think about interface, interactivity, and meaning/readability. In the latter

two-thirds of this article, I explore what interface and interaction means in each field, where the concepts overlap and how they differ in the two fields.

To foreshadow my conclusion, I argue that architects work with a more embodied and social notion of humans than HCI. HCI might do well to consider what is missing from their models of humans if it is true that they do take a less embodied approach. We need to explicitly ask deeper questions about what is special about the human side of the human–computer relation. Likewise, foundational questions about interface and interactivity have also to be asked because when we design for humans we want to be informed by our best scientific theories of what it means for humans to interact. These fundamental concepts shape a field. By clarifying and systematizing these ideas, I hope to open a conversation that will lead to reframing existing views in both HCI and Architecture.[1]

A few words about my approach. For several years, I have been resident at the Bartlett School of Architecture trying to understand the distinctive elements of architectural thought. I have had the help of many architects who explained their views on dozens of questions over scores of hours, related to their ways of thinking about buildings and people's relations to them. My goal here is to present ideas that architects often seem to assume without explicit discussion. If the result appears to border on the philosophical, it is because the major theme of this inquiry is how embodiment plays out in the foundations of architectural thinking and how this can inform HCI. The points I make about interface, interactivity, and architectural readability are not ones to be found in books or architectural articles; they are the outcome of my effort, as an outsider, to make sense of how architects speak. The result is a somewhat novel theory of interface and interactivity and a view of architecture that has my own cognitive bias. In the end, many of the differences between HCI and architecture stem from the undisguisedly embodied and social nature of architecture. Social interaction is between flesh-based humans in proximity. Although virtual and digitally mediated beings are becoming more sophisticated and common in buildings, and we soon may share our activity with digitally clothed colleagues, the soul of architecture remains tied to space consuming proximate humans. People we can shake hands with and embrace. What this means for HCI and Architecture is inevitably a bit philosophical.

What follows next are what I believe are foundational differences in the way architects and HCI practitioners think about design. Those impatient to see how the two fields think differently about interfaces and interactivity may skip to part two or part three.

2 PART 1: DIFFERENCES IN THINKING BETWEEN HCI AND ARCHITECTURE

Here are six ways architects and practitioners of a somewhat idealized form of HCI think differently about their design problems. Of course, HCI as practiced today, is broad with specialists working on topics that look very different than the old days of designing interfaces for computers. Nonetheless, I believe a vestigial mindset remains that keeps architects and HCI designers from fully understanding the other's perspective. This difference in mindset also represents a challenge I personally had to grapple with in order to see things the way my architectural informants do.

2.1 Social vs. Task Focus

How would you design a room like that in Figure 1: room shape, lighting, windows, ceiling, and passageways? To design a space, an architect needs to know who will be there, how long, what they are likely to do, how they get there and how they leave. Architects design for activities of all sorts, from manufacturing and assembly, to professional cooking, working in foundries and

---

[1]After completion of this article the author found a provocative earlier effort to jump start a dialogue between architecture and HCI. See [17].

Fig. 1. A space designed for different forms of social interaction and work. Image credit: Gensler Microsoft London.

offices, and taking enjoyment in retail shopping. Still, it is often said that the heart of building design is about supporting social activity rather than task performance per se. In virtually every functional context, social encounter is important. It is a tenet of modern design that much of value creation comes through meaningful social exchange and team work. "In Silicon Valley the tight correlation between personal interactions, performance, and innovation is an article of faith" [84]. Architects are expected to design to facilitate personal interaction. They must focus on how space and surfaces can be structured to help people interact. This holds as much for shopping, as for office work, individual creative work, and even production lines, where observers and managers must pass through.

What is special about social interaction? In physical space, it is typically face-to-face, or within speaking distance, and it is embodied in that it takes place between space-occupying agents with faces and backs and shared situation and social awareness. When two people are in the same space, they share knowledge of what is around them, they can see where each is pointing [80], they can act jointly on nearby things, and they can engage in full body interaction – what some have called performative actions [33]. These are the starting assumptions of architects.

In much of HCI, social interaction is important too. Yet, historically, the standard interface was designed for single users working with digital input and output [61]. Requirements were based on functionality and task needs of that individual at her station. In Computer-supported cooperative work (CSCW), where the focus is explicitly on collaboration and social interaction, the mindset is group oriented and indeed joint activity is fundamental. Central questions, though, are less about physical presence, and embodied social interaction than on sharing digital information and coordinating computer-mediated activity. The metrics that matter most in CSCW have been (and remain) related to task completion and effectiveness [59]. Even context-aware systems operating ubiquitously and invisibly still have as objectives the notion of delivering a functionally specified

service [25]. Architects, perhaps to their loss, rarely have specifications that involve such task or service related metrics[2] because their focus is more on the embodied and social use of space.

2.2 Immersive

Buildings are essentially 3D forms and much larger than humans. Experience suggests that the "spatio-temporally immersive" experiences that users have in buildings are very different than users' experience of artifacts that are manually controlled [2]. Working on something is different than being in some place or working in there. When interaction is hand sized, it takes place in peripersonal space – the space within hand and arm's reach [72]. This certainly applies to hand-sized input devices and monitors whose WYSIWYG style reinforces the sense of hand control. Things have changed in HCI. But the imprint of peripersonal thinking is evident even in today's more contextual HCI. The inspiration of direct manipulation is strong.

In buildings, interaction tends to be body-sized, taking place in *extrapersonal* space, where people interact among themselves and move about [81]. Navigation is not like direct manipulation; nor is social interaction. Naturally, there are moments when people are touching knobs, handles, and switches – situations where interaction is peripersonal – and at such times architects must think about direct manipulation. But supporting how people act in open space requires a different model of interaction.

Along with the neuropsychological differences between acting in peri- and extrapersonal space [30], there is a cognitive difference related to perspective that comes when a person adopts a first person point of view (POV) when acting [12]. When objects are small relative to us, they can be seen from above, and from multiple angles. In principle, we could manipulate them in our peripersonal workplace where we can see them from all angles by rotating them. Monitors bolster this feeling of acting in a manual workspace. When things are large relative to us, however – big pieces of furniture, or the spaces we traverse – we cannot survey them as if from heaven. Think of the difference between playing with a doll house where one can see all corridors and spaces from above and we are able to move "large" furniture by hand vs. moving real sofas and walking through open space from room to room. Or consider the difference between playing with a toy car on the floor and driving on the highway. When we play with small things, we assume a more cosmic, god's eye perspective that includes a feeling of control we do not have inside real traffic or real buildings [14]. David Marr captured the spatial portion of this idea by distinguishing object-centered representations – where we conceptualize an object as a 3D form rotatable in any dimension (as if in a 3D graphics system) – from subject-centered representations, his 2½D representations that are shaped 2D surfaces, viewed from a point in 3D space, like the skin on our face viewed from a specified point in the room [53, 54]. Architects think about interactivity and interfaces as being experienced in a world of subject-centered 2½D surfaces and social encounter but created from countless points in 3D geometry. What does it look like from this angle, and this perspective? See Figure 2. People inhabiting space always have a POV. People handling objects

---

[2]The absence of metrics for the efficiency and effectiveness of a designed space has often been held as a sign that architecture has yet to reach the evidence-based stage. Part of this challenge reflects the credit assignment problem: how well a building is working for its occupants measured by the success of activities depends not only on design but on work practices, technology, and furniture, all acting over time, making it hard to assign credit or blame to any one component or any one time period. The result: almost all post-occupancy studies confine their tests and evaluations to light and air quality, energy efficiency, and other physical design parameters that may affect workers and building performance even though these effects may be small compared with the impact of structural layout, interior design, architectural beauty, novelty, not to mention team composition, work practices, social happiness, and so on. Without a good measure of occupant efficiency, effectiveness, and happiness on the one hand, and a good method of describing the meaningful structural differences of spaces on the other, architecture remains an evaluation challenged design field.

Fig. 2. A staircase in the Villa Savoye, Poissy, France designed by Le Corbusier and his cousin, Pierre Jeanneret, 1928–1931, is a perfect example of how architects think of structure from a 'multi-point first-person point of view. Here, we see how Le Corbusier used black hand rails to create Mondrian-like limned forms visible from certain angles. Image Credit: europaconcorsi.com.

can take a more 3D view, conceptualizing the whole shape almost at once because manual objects can be so easily rotated.

The sooner HCI's God's eye approach is swapped for a "multi-point" first-person POV, the more relevant an architectural perspective becomes to HCI design. It has already happened. The HCI that increasingly matters now is woven into the fabric of our everyday lives. Interaction is moving fast from dedicated input devices, like the mouse, joystick, wand, mobile screen, and watch, which are all hand sized and under a user's intentional and manual control, to devices that are less visible and more appropriate to inhabiting space, such as ambient, sensor-based technologies that are in-body, on-body, or distributed throughout our environments. Interaction is becoming more embodied and tangible [32]. More implicit too [37]. Understanding new notions of interaction and embodiment is bound to be useful as HCI pushes toward a more inhabiting, immersive model. Architecture might offer some lessons. Certainly, notes should be compared.

2.3 Unmediated

When two people collaborate, much of their time is spent in unmediated interaction; just doing things together. There is no interface boundary between people, it is a social relation. When people interact with digital systems, by contrast, their actions must be mediated in order to cross the digital–physical boundary. Because feedback from the digital side is fast and often perceptually realistic, it engenders a sense of direct manipulation and agency – a feeling that interaction and control is unmediated. Nonetheless, control is always mediated because we use tools, input devices, to effect change. Even when the input device is a camera and there is no special action we must perform to create a signal, it is still clear where the input surface is. It is the camera lens. It constitutes a boundary that must be crossed.

There are further differences between interacting socially and interacting through a digital interface. First, when people interact, they choose where and how: with coffee nearby, a table between them, while sitting on chairs, or a couch, with a whiteboard in easy reach or while they jointly look at their phones. These physical elements are part of the social interface they inventively

Mediated, Less Mediated, Unmediated Interaction

| Figure 3a | Figure 3b | Figure 3c | Figure 3d |

Fig. 3. (a) There is a clear physical interface between hand and screwdriver, screwdriver and wood. A tool mediates our actions on screws. (b) A digital mouse is a digital–physical interface point. By moving a mouse, we interact with objects represented by the images on the screen. (c) When two people interact face to face, it is less apparent where their interface is. Spoken language and gesture are obvious mediators and might be said to provide an interface that both parties use as tools to pass thought and express feelings. But that view is limited because proxemics, the shape of space and one's surroundings all affect how people interact and what their words and gestures mean. Nearby things are props and shareable referents that challenge the simple idea of mediation. (d) What mediates two people shaking hands? Their joint hand action constitutes the process of shaking hands. By definition it is unmediated.

create. If we ask where the interface is, given that there is no set of physical things that must serve to transmit signals, or that are required to mediate social exchange, it is likely to be found in the comportment of people and their changing relation to physical objects and others in the space. For instance, chairs and coffee function like props on a stage; they help create the setting in which people interact. But they don't mediate social interaction the way mouse and touchscreen mediate human–computer interaction; they scaffold it. There is no physical–digital boundary between people. "Props" and their layout are part of the live context where people make sense of each other, determining the kind of person the other is, what their demeanor means, their level of engagement. This happens through embodied presence in shared physical space. The objects in that space can help to shape the process of social interaction [7] such as when a person moves a chair. But these objects don't create a boundary-like interface that must be crossed for interaction to take place. See Figure 3.

Second, when people communicate via speech and gesture, the semantics of these communicative media is far more complicated than the semantics of input devices, where signal and interpretation is predetermined. It is tempting to suppose that social interaction is mediated by language in a precise way, and that this mediation is analogous to sending a signal in a fixed code through an input channel. As many years of research has shown this downplays the importance of context in interpretation. The meaning of an interpersonal action, even a speech action, is highly contextual and relies on people sharing an understanding of situation and context. Speech is situated and indexical. No simple code or input medium completely mediates this shared understanding. It is built up, negotiated, and dependent on dozens of social cues. It relies on shared understanding and meaning making.

Accordingly, people in buildings communicate and share action quite differently than humans do with computers. Sometimes an object – e.g., a seesaw between children – does indeed mediate activity. But what mediates the joint activity of shaking hands? Hands are not tools. In HCI systems, mediation is required for interaction. In social interaction, it is not required, and when props and tools are present their mediating role is far more complex than just carrying signals. This difference runs deep.

Fig. 4. An extreme example of appropriating an object for a purpose it was never designed for is seen in Duchamp's famous urinal displayed in 1917. This became known as found or appropriated art.

2.4 State Space vs. Social Space

Architectural space is not usefully conceptualized as an environment of forced choices – i.e., a network of choice points where people must choose among a small set of possible ways of occupying a space or doing things in the space. In software-driven systems, even in networked and distributed systems, the functionality of tool, apps, sensors, and actuators is well specified in advance. In classical HCI, specification is even narrower: it is the space of physical-inputs and digital-outputs defining the use of tools, buttons, sliders, and filters. Even in invisible context-aware systems services are well specified. This means that when interacting with digital systems our actions are not just constrained, they are state space driven. Where we have choice, our decisions come down to choosing this or that parameterized action. This has the effect that we can't use a tool in a completely novel, unprogrammed manner.

How different our everyday world is. In addition to our freedom in speech, gesture, and social action, we also are free in how we use the objects around us. If the urge takes us, we can throw pillows from a couch to the ground where they are not meant to be placed and use them there as backs, seats or bolsters; we can pile them on stairs and make mock chairs; we can take hot coffee and threaten someone with it, or use it as a prop in a discussion. Although some architectural spaces have preferred uses – kitchens, bathrooms, studies, and TV rooms – people still co-opt space relentlessly. We *inhabit* space, and we *appropriate* the artifacts around us for our social and practical ends [9]. See Figure 4. When one inhabits and appropriates, the relation between body, space, cause, and effect is quite special. It is situated, embodied, embedded, distributed, enactive, and often extended [41]. We can be inventive and creative. The result is that Markov models and AI state space models are descriptively inadequate because those formalisms require that all options must be represented in advance and we do not know the full range of options [34].

This freedom to change things and innovate is only weakly duplicated by reconfiguration and personalizing in digital systems. Change is lightweight in the social world. It's nothing to reposition a chair, leave a door open, and alter the lighting. Our spaces have been designed with reconfiguration in mind. Physical space may be inhabited in multiple ways in the course of a single day. Consider how a breakfast nook in a shared house might be used: for eating, reading a newspaper, private studying, intimate discussion, repacking a backpack, and working on a computer. It is a place to do many different things, and often each activity requires laying things out and moving things around. Reuse of space is intuitive not contrived.

Because alteration of physical space is lightweight, it happens far more frequently than personalization or customization in HCI [8]. And it is collaborative. In buildings, people regularly tailor make their interpersonal interfaces. They move chairs to face each other or to face sideways to avoid the feeling of confrontation. They reposition a chair to rest their foot on a ledge. Such small

changes can have outsize effects on the success of a social encounter, suggesting intimacy rather than formality. Owing to the centrality of social activity in buildings, some architects think of an interface as a 3D spatial ensemble, the volumes and forms that shape a communicative context.[3] These sorts of interfaces come and go. They are configured by the participants in the short run, but there is no doubt that initial room design plays a major role. And there is no doubt that architects have views about how different spaces should be architecturally arranged to support the *customary* unfolding of social activity.

For instance, it is accepted wisdom that in most buildings there is an "intimacy gradient" [3]. A personal powder room shouldn't be by the front door, or the public space for guests right at the back of the house, forcing them to make their way through intimate areas. These ideas are present in architectural thinking when decisions about ceiling height, room size, and amount of natural light must be made – all factors that affect our sense of seclusion, quietness, and privacy.

Once we accept that there are *emergent interfaces* – places of social encounter – that arise because of how architecture, interior design, and the momentary desires of occupants meet we have to wonder whether the coming wave of AI-driven context-aware systems will operate with a rich enough model of users. Architects have learned to balance their ideas about how space should be used with an appreciation that occupants reconfigure and repurpose.

2.5 Procedural Complexity

HCI arose from the need to tame complexity skillfully [77]. Complex systems like power plants or plane cockpits needed to be controlled more reliably. In application environments, the problem is often to stay in control when many steps and sub-goals are required for task performance. Almost nothing we do in buildings is like this. A person may, at times, have to stop and think about the steps needed to rearrange furniture; or they may take 5 minutes to create complex lighting that requires sequencing of switches. But these are exceptions, not the rule. Cooking can be complicated; its complexity, though, is not rightly seen as architectural, since so many of the parts interacted with are pots and pans brought into architectural space. Running machinery in factories is complex, but it is not part of the architecture. Everyone agrees that architects play an important role in designing the space for machines to operate and pass products. Designing a production line or a robotic playpen regularly requires architectural involvement. But the job of an architect stops before the design of the controller interface.

So when does an architect think about facilitating long chains of interaction? To date, aside from rooms for controlling the building, not often. Again, the reason is not that people don't do complex things in buildings; it is that they do not interact with a building or any of its niches in such complex ways. That is not what building interaction or building interfaces are about. So far! It is one explanation why the fields have historically had such separate concerns and training, and a major reason why architecture will increasingly need HCI as greater digital interpenetration of the physical raises the complexity of building interaction.

2.6 Sense Making

People don't read buildings the way they do HCI interfaces. Each medium draws on different metaphors and narrative inspirations. This idea of meaning making goes beyond the semiotics of a given design where the symbolic meaning of a design is front and center. In Figure 5, Frank Gehry's Binocular building made with the sculptor Klaus Oldenburg trumpets the idea of surveillance, seeing into the distance, being a visionary. Perhaps, it is fitting that it is now owned by Google. This is building semiotics.

---

[3]Several of my informants promoted this view.

Fig. 5. The Binocular building in Los Angeles by Gehry and Oldenburg designed for the ad agency Chiat/Day is a design driven by semiotics. It is more sculptural than functional. In this case, the exterior is more concerned with narrative than efficiency. Photo courtesy of wallyg, Flickr.

There is another sense of reading a building, however, that is deeper and less showy. Architects are always looking for a good story to explain the design decisions they make about a building's morphology.[4] In architecture class, the narrative that students tell about their design is often as important as the design itself. What idea does the building implement? What was the guiding thought? One reason that buildings often incorporate a narrative is that the design space of architecture, as any architecture professor will tell you, is so large that architects must give themselves radical constraints to make their task doable [11]. This is a point my architectural informants stressed. They need a guiding idea or a few distinctive features around which to create their structure. This doesn't usually apply to designing a small extension to a residential building, or to redesigning a kitchen, where efficiency and functionality, as in HCI, drive imagination. In designing larger structures, the architectural question is less about efficiency and functionality (primarily); it is more about narrative and making a space meaningful and special for people.

The design space in HCI is large but orders smaller than in architecture. Narrative is important in product design, at times, but the scope of the narrative is also smaller, the ambition less grand. So far, people do not read HCI interfaces the way they read buildings.

Still, there are important overlaps. Part of the idea of making a space meaningful is indeed similar to the HCI notion of affordance and making an environment readable. People need to be able to see what they can do in a space: Which part of an HCI space, such as that defined by a

_____

[4]For a beautiful example of a building's narrative, see [49].

'window', is workspace and which parts are for displaying digital tools? What can those tools do and what must you do with them to have them execute their function? This looks like reading a church, where one can see where to expect the sermon to be preached, or the choir to stand. Or how one experiences a restaurant with its booths, bar stools and interior tables. This is prosaic reading off of function, but it is an element that is between the two.

This completes my account of six areas where there are differences in the way architects and HCI practitioners frame their problems. The two groups come at their problems differently, with different design goals, values, and presuppositions. Context-aware computing is bound to change this. The question is: will it change HCI into something more like architecture, with its very different notions of interface, sense making and interactivity? Or will it go its own way? It is to these big questions I now turn.

In part two, I first articulate what I believe is a pervasive but implicit view among architects of what an interface is. I then contrast it with two concepts of interface, then I return to complete the architectural view of an interface. The implications of those concepts for interactivity are discussed in Part 3, where three types of interactivity are discussed.

3 PART TWO: INTERFACES

3.1 Architecture

In architecture, the interface of greatest interest is an emergent structure. It is not predetermined as in HCI with its well-defined input and output devices. Rather, it arises from the way the volumes, surfaces, material, furniture, colors and, above all, the location and nature of the people who occupy a space dynamically determine the structure of their interaction. How? By determining who and what can be seen, heard, touched, what can be used in joint activity, what is in shared peripersonal and extrapersonal space and what can be assumed to be shared epistemic state. When people choose where to sit and chat, how close to sit beside each other, the type of lighting, they co-create an interface they are an essential part of. People constitute part of the interface for each other. This is nothing like the concept of an interface in classical HCI or even in networked, distributed AI-based HCI where an interface is where inputs are captured, and outputs distributed. It is an embodied and ecological view incorporating the ideas of joint activity theory and social co-presence.

A striking example of how architecture partly shapes the interface or social niche in which people interact is seen in the way a serving hatch constrains social interaction between kitchen and dining room in a 1960's suburban American house [22].

A hatch, also known in French as a *passe-plat*, narrows the space of interaction between those who are working in the kitchen – typically the wife in that era – and the husband and guests in the dining or drinks area. See Figure 6(a). The hatch obviously structures the passing of food and drink, but it also structures who and what can be seen from each area – that is, the situational awareness of each participant; it determines how well sound travels and, accordingly, the degree of full social exchange that can take place between the areas. One would not expect a serious exchange on the meaning of life through a hatch, nor a private tête-à-tête. In Figure 6(b), the hatch now includes a place to sit and drink or snack. In deciding to add a hatch to a house, an architect enshrines a set of social conventions concerning who interacts with whom, when, and how they interact – i.e., social roles. This architectural element thus defines a social, epistemic, and functional interface. Once a cast of people populate the space the interface is complete and more precisely defined. Figure 6 shows how varying the design of a hatch and cooking area defines a social interface.

Materials are another component of an architectural interface. Transparent glass creates an interface surface where events on each side can be seen but not heard; steel creates another. When a wall is opaque and soundproof, the things on the other side are outside the interface. Floor

| **Figure 6a** | **Figure 6b** | **Figure 6c** |

Fig. 6. (a) A serving hatch partially defines the social interfaces afforded by this space. It carries meaning too, in terms of who may be expected to be in the kitchen, the social roles they play, and the scope and tenor of conversation. (b) The hatch now includes a place to sit and drink or snack. The social relationship between cook and diner is closer because stools allow a sushi bar style relationship. (c) The kitchen is completely open to the main room. Food preparation is essentially a social activity, a show where guests watch or even participate. Conversation is unconstrained.

Fig. 7. A room like this offers the possibility of many forms of social interaction: conversation, cleaning, play, and sharing the closet. It contains many potential interfaces. Each comes into being when people with certain relations animate the space. It is like a system of multiple overlapping niches in ecology.

materials like carpet, tile, or marble also affect the interface. Each material affects acoustics, line of sight, ease of movement, comfort, privacy, how things feel – all attributes that help shape social interaction. "God is in the details."

   None of these effects are deterministic. Look at the room in Figure 7. The space, surfaces, and furniture define a collection of possible interfaces. In the same room, there is not one possible interface; there are many, depending on who is in the room, the social conventions, the activity the participants are engaged in, and the part of the room they occupy. Each interface, once created by people in place, shapes the actions that are possible or invited and the actions that are denied or inhibited. In the spirit of Simon and Newell [62], one might try representing these interfaces as a set of task environments: one for tidying up (if the participants are cooperatively cleaning up), another for dressing or chatting, or playing on the computer. Each task environment would define a set of fixed choice points and option sets. When immersed in a task, subjects, according to this model, would choose one among the task relevant options as they move from one choice point to another.

As valuable as this perspective has been in studying decision making and problem solving in some domains, such as games and well-defined problems [48], in architecture, it seems utterly forced. Imagine two people, one sitting on the bed, the other sitting at the desk. How many different kinds of actions are possible – including speech actions and behavioral surprises? The option set is ill-defined [78]. No architect would try to impose task environment formalism. It is too constrictive. Still, everyone knows that not just anything can happen in a space. Over time it is quite clear what the distribution of different types of behavior is. Clearly, the environment promotes and inhibits certain actions. The twist is that the people who occupy a space co-create and redefine the actions it affords, so it is not just tasks that matter, it is the setup and the sort of interactions that the participants tacitly agree are acceptable. This tacit negotiation of acceptable or plausible activity is the respect in which social interfaces are organic, like biological niches. The material properties of the niche initially constrain activity, but things often change in the course of interaction. The implication: interaction can change the interface!

The idea that interaction might change an interface runs contrary to HCI conceptions where the point of an interface is to define where agent and system meet and where each may have effects on the other. The interface itself is not changed because it is already predetermined. In biology, however, a niche is a dynamic notion that can and usually is changed by the population that occupies it. For instance, cows both fertilize and dry out the land they live on as their population grows. When a new organism enters the "same" terrain, the other organisms adapt to accommodate the "intruder" and each niche realigns as realities change and they all dynamically rejig what they eat and how they live [69]. Organisms co-create their niche. Hence, we cannot define a niche independently of the agent or agents who inhabit the niche and the effects of their behavior over time. It emerges; it cannot be predetermined.

In human habitats, when one group moves furniture their niche changes subtly. This means that we cannot fully understand the interface that a group creates – the niche it creates – unless we understand what the physical setup is, who the group is and the tasks or activities they do in that space. We need to know their cultural predispositions, their etiquette, and so on. Architects design with people in mind; they have ideas about the types of interface these people are likely to want and create. An architect's job is to design the flexible substrate that supports these social and activity niches – the spatio-socio-technological interfaces that people co-create. See Figure 8.

In HCI, the model of an interface and the interaction it supports has changed as the field has advanced [see 35, 36, 31]. In the next section, I distinguish two models: a model typical of classical HCI also applicable to tool use more generally, and a model that fits context-aware HCI, AI-based HCI and distributed systems HCI. Both are significantly different than the niche model of interface that I suggest is the driver of much architectural thinking.

3.2 Direct Manipulation Interface: Classical HCI

In the early days of HCI, an interface was understood as the boundary between physical and digital systems – where two systems meet. The original idea was that an interface is the set of I–O channels through which a person A and a digital system B communicate and interact. The two abut at a physical/functional surface, where user activity on input devices crosses over into the digital world.

This boundary is most naturally thought of as an n-dimensional I–O surface. Each dimension is an information channel – an input or output stream – each with its defined capacity and expressive power [13]. On the human side, a person acts on known input devices that transduce physical activity into electrical signals that cross the computer boundary and are then collapsed into a small finite language of discrete impulses that represents the signal. Interpretation of signals proceeds through a chain of programming contexts. For example, moving a mouse may be interpreted initially as
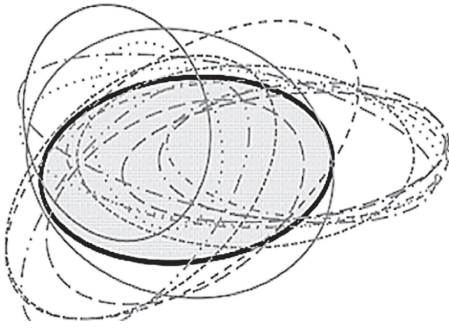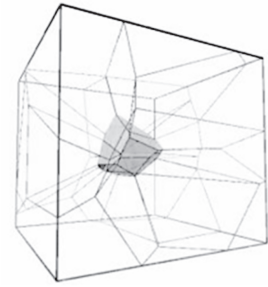
Figure 8a

Figure 8b

Fig. 8. (a) Represents an ecological niche that dynamically reshapes because of the effects the inhabitants themselves have on their terrain and food and the arrival or departure of other species whose niches spatially overlap. (b) Represents the dependency of a niche on neighboring species. The niche of one species is constrained by the niches of other species who live nearby. It is not predetermined. These type of dynamic models of how a niche changes better fit the notion of an architectural interface where the social bubble people construct depends on how many are in it, who else is around, who enters the bubble's space, where things are positioned in the bubble and what's nearby, as in an open office layout. Removing a table or chair, or opening a door, would change the socio-spatial interface, as would introducing a mobile phone or a group printer or water cooler. Such a niche remains a social construct even if there is only one person in it, though at such times the task and physical layout are the dominate constraints of the niche.
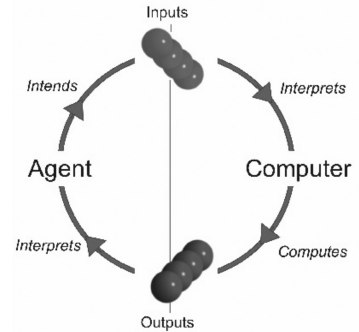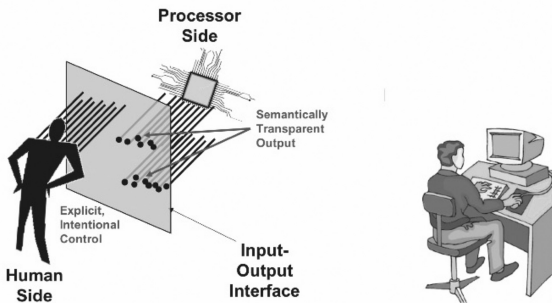

Fig. 9. In the classical view, an interface is an n-dimensional surface where inputs enter through one set of dimensions and outputs exit through another set. The agent or user drives the system by choosing inputs. The system displays feedback on the input actions as well as program results. User and system behave as a closed loop.

moving the cursor from position (x_1, y_1) on the screen to position (x_2, y_2), at that point, it might mean grabbing a corner of a geometric cube and rotating it, which is represented internally as a cube of a certain size and color, relative position in a 3D space, and so on. Likewise, a keyboard tap on lowercase T might be interpreted to mean add "t" to this text field or assign value "a" to a variable, depending on programming context. Output from the digital side emerges as a change in the visual or auditory channel of the display, or a page being physically printed, and so on. See Figure 9.

Nowhere in this old HCI model, is there room to think of *where* the agent is. The relevance of his sitting in a chair or lying in a bed is not part of this type of HCI interface. There is no need to characterize the *presence* – social, physical, and epistemic – of anyone because people are a

construct defined over the set of predetermined actions that they can submit through the input channels. They could be robots or remote controllers modulating mouse and other inputs. Effectively, a person is a function that takes their own goals and objectives and maps these into actions to be performed on input devices in light of the output and feedback provided by the digital system. Person and computer therefore form a loop: the user intentionally acts on input devices according to goals, the system reacts according to its operating system and active programs, including providing feedback on display or audio channels. The cycle recurs.

Architects, on the other hand, quite naturally care about where a person might be and the sort of social, physical, and epistemic presence they may have. Architecture is about embodied interaction, between sentient, space filling beings. When a person is viewed as a goal directed system independent of bodily presence, then their location, inner feelings, social, and other non-input interactions become irrelevant.

Another assumption of this very classical interface model is that agents know what things are input devices, how they are to be used, what their input actions mean – as long as feedback is well-designed and revealing. They therefore act on these devices intentionally and users *feel* they are in control. Input devices are typically hand sized. This hand-based control paradigm of *I do – it does* led the community to see interaction sequences in classical HCI (really interaction patterns) as akin to dialog where turn taking is required. Another metaphor was to see interaction as akin to working with objects or tools, as in assembly or sketching, where agents can directly manipulate things to cause change. Richer interfaces soon supported more types of actions, each requiring a shift in metaphor – queries, browsing, complex tool use, controlling the parameters of a process, and visualizing. Nonetheless, the direction of interaction continued to be primarily from user to system. As the complexity of tasks increased the metaphor often changed from dialog to one of support. When a task requires managing many complicated steps, where some of these steps are dependent on completing others and sub-goal interactions become complicated, a thoughtful interface can help the user by maintaining visibility of system and task state, it might offer recommendations, provide buffers that store interim results. In such situations, the functionality of interfaces goes well beyond dialog to something closer to epistemic management. Nevertheless, despite the need to change interaction metaphors, the basic structure of such an interface still seems most naturally understood as a boundary where the user intentionally controls things via input actions s/he understands and is aware of using; the system responds with feedback, recommendations, and operational performance from its side. The feeling of control is important.

A key step in the phenomenological feeling of being in control is that users move beyond registering a boundary between system and human [19, 57]. This is well known when applied to tools e.g., screwdrivers, hammers, pencils, but it also applies to classical interfaces.

To dissolve a computer boundary, two phenomenological conditions must be satisfied: (a) control transparency – a computer user must feel that her input devices, her mouse, and keyboard, are "transparent" – she *sees through them* and *controls through* them to operate "directly" on their onscreen counterparts; and (b) semantic transparency – users see beyond the immediate onscreen effects of their actions and take a second leap of intuition by seeing through onscreen icons to their meaning.

Here are the two components in detail:

(a) Transparent control – feeling in control: Input devices feel like gloves or glasses – they are things a user perceives through, not things they perceive. This happens when there is a sufficiently tight coupling with the system that users feel their actions are operating on the onscreen digital entities directly – the icons – rather than on a mouse or screen. The analogy usually given is with a blind person's white cane: its distinct physical presence as a

cane dissolves as it is absorbed into the sensory system of the acting person [55]. The tip's contact with the ground is what is sensed, not the mass of the cane or its inertia. Just as we rarely feel our eyes move when we saccade or look sideways, so a cane user rarely feels her cane move because her attention is on the forces at the tip, and through those to features of the terrain. The target of attention lies beyond the physical cane.

(b) Transparent semantics – interpreting screen events semantically: Onscreen entities are intuitively meaningful, so users see past the icons or letters to their referents or meaning. When reading a book, no one looks at *words* as *letter sequences* like we do when looking at non-words such as uhsbxn mckdjg. We see in semantic units. Likewise, a user, call her Jane, doesn't think she is dropping an icon of a file *onto* an icon of a folder; she thinks she is dropping a real digital file *into* a real digital folder. It feels to her as if her mouse operates not just on a curser but on a functioning curser, one that lets her act on application objects. This is possible, in part, because classical HCI systems are closed. Exogenous inputs are not supposed to occur. Jane is in control. So when she drops the icon correctly, it does what it should. Likewise, outputs are supposed to be rule determined. These deterministic rules make it easy to learn the semantic function that maps screen entities to their referents.

We can summarize the normative model of a classical interface as a human–system relation that satisfies the conditions shown in Table 1.

These conditions are not met in architecture where people are an intrinsic part of their interface. Other than our individual interaction with knobs and switches etc., where a direct manipulation model clearly does apply, there is no discernable n-dimensional interface in architecture with channels, bandwidth, and signal meanings. Because we never think of other people as input devices,[5] it makes no sense to think of an interface in architecture as a place where one side, a person, meets another side, the thing controlled. Social interaction is not like that. For example, when people dance the distinction between input and output is blurred because it is a joint activity without clear directionality at each moment; control can move back and forth, or be truly joint. Or when people use a nearby object as a prop in a conversation, they are not using it as an input device; they jointly breathe meaning into it. So, it is absorbed into the social bubble where meaning is situationally created.

A third respect in which architectural interfaces differ is that not all interface actions are intentional and explicit. This follows because in social interaction some interactions are implicit: body stance, orientation, mutual distance, and facial gesture [23]. In virtue of having a body, participants always have a location, orientation, relation to other people, and building surfaces, whether they want to or not. People can't help but generate "input" in the sense of doing things that carry meaning for others. These "inputs" are often unintentional, implicit, and non-transparent. They are not part of the direct manipulation paradigm where agents feel they are the cause of events across the interface boundary.

Last, architectural interfaces are not closed, deterministic, and independent. The interface changes as people move about, as more enter, as new tools or props are brought in – all ele-

---

[5]Ears and eyes might be thought of as input devices, but importantly we do not control the full signal entering these. The senses of others are not direct manipulation input devices. Moreover, were we to start thinking in this mechanistic way the range of input devices we would have to say a person incorporates is broad and not semantically transparent. If I stroke, shove, trip someone, for instance, how many channels of their input "system" have I acted on? All have effects, but some effects are non-informational. The ones that are informational, such as words and gestures, usually do not provide transparent feedback from the receiver to actor, echoing the meaning of the signal. We don't tend to repeat the message or say: "copy that." Moreover, our words are ambiguous, highly contextual, and their meaning or impact on a listener is a complex function of many things. Accordingly, the degree of control agent A has over input signals to B falls far short of the requirements for direct manipulation.

Table 1. Attributes of a Classical Interface

Attribute	What each attribute means in classical HCI – Direct manipulation
Recognizable input devices	Humans know what and where all the input devices are.
Manual control	Input devices are typically hand sized, and directly manipulable.
Feeling of control	Humans feel they are in control.
Explicit understanding	They explicitly know what their actions mean to the system, at least superficially. There are learnable semantic rules linking digital images onscreen to digital entities deep inside the system.
Intentional control	Humans intentionally control input according to their goals and their understanding of the meaning of actions. They are the initiators.
Closed	The entire I–O system is closed to outside factors. Only the user can cause changes in input devices and only the digital system can cause changes in output modalities.
Circular	Users can be understood as a function that maps system outputs to inputs to the system in a goal-directed manner. Changes in the ambient context of a user do not change how an optimal user would behave given their goals and beliefs and so a system need not monitor more than the signals a user provides via input devices. The low-level meaning of input is not affected by external context.
Deterministic	The system does not respond in an indeterminate or stochastic manner. It is programmed to respond to inputs in a pre-determined manner. (This also includes programs that are designed to have a random response as long as it is pre-determined.)
Independence	Users are not part of the interface, they interact with it. Hence interactions with the interface do not change it (in the short run).
Direction of adaptation	Humans must acquire the skills necessary to use input devices appropriately. It is humans who adapt to interfaces rather than the digital systems, though of course good design makes it easier for people to master the system and systems may make small peripheral adaptations to users. See figure 10.

ments that violate closure. Actions that occupants perform change the interface – violating independence. And people are unpredictable, so in social interfaces input–output functions appear to have non-deterministic elements.

3.3 Network Interfaces: Context-Aware/Ubiquitous Systems

The differences between HCI and architectural interfaces partially fade when the system a person interacts with, is context aware. Because the overall system is distributed, partly invisible, and system output may change the participants' activity space it may seem to us as if we are not

Fig. 10. Humans adapt themselves to classical HCI interfaces when they master the skills necessary to control digital events via input tools. Image credit: http://www.kellychiropractic.co.nz/.

on completely different sides. The boundaries seem to blur further when the "system" contains multiple people distributed remotely, all of whom interact via local devices networked together. Now, the "HCI" system no longer *feels* closed and deterministic. It does not feel closed because any one person's actions may not have its expected effects. If two of us speak at the same time, or attempt to act on the same shared object, we cannot predict how the system will behave. Prediction fails not just because algorithms are too complex – though they often are; it fails because of the impossibility of knowing the timing of signals. We cannot predict how a network will react owing to the unpredictability of who will do what, when, what traffic there is on the net, and how dynamic adjustments are made through adaptive programming. What happens next inevitably depends on moment to moment exogenous factors that are unpredictable. Hence, the system feels open and non-deterministic because it partly is. Actions on the system have probabilistic outcomes. This is one profound difference between old fashioned HCI and distributed HCI. It does not technically dissolve the boundary between physical and digital, but it feels as if it is beginning to.

Another difference between classical and pervasive, ubiquitous or context-aware systems arises from passive sensors. Once sensors become an important input source, users no longer have explicit knowledge of what their actions mean to the system. Do I know what a Kinect[TM] represents me as doing when I pull a long espresso? Kinects often compute stick figure movement based on infrared motion and depth detection. Other AI-based vision systems use additional methods. Do I know how to act to create input that will lead such systems to represent my current action correctly? For instance, when I want a system to know that I am filling the portafilter with ground coffee what must I do? And what should I assume it knows? My grasp of the semantics of its input representations is incomplete. And possibly, my grasp is gappy because not only am I uncertain what the sensors are capturing (since I may never see their output) I may also be unaware of how many sensors there are, where they are, and what they are doing right now. Sensors can "manipulate" themselves – auto-on or auto-off. Do I know which are on now and which are off? This further weakens my sense of agency and threatens my capacity to control input explicitly and intentionally. In fact, who initiates interaction, me or it?

This radically changes the idea of an interface. If people are unaware of pervasive sensing, why should the system offer them feedback of what it is tracking? Showing such content might be distracting, like watching oneself walk upstairs.[6] Worse, what the system knows or derives from

---

[6]In retail stores, where continuous camera feeds are broadcast on visible monitors, the objective of the display is not to tell occupants what they have just done; it is to remind them that they are being recorded. This is quite a different purpose than providing feedback to facilitate transparent control.

sensor input is opaque to the user, even if the raw capture were typical video and not infrared motion. It is opaque because video input must be interpreted. Why suppose that a camera represents the current situation in the way *we* see it, even when it computes stick figure movement, or that the system collecting video categorizes things in the same vocabulary we use to characterize actions, beliefs, expectations, and predictions? Its classifications may be incomprehensible to humans. This certainly would be the case if the system relied on a classification of camera input that was derived from machine learning of big data. Meanwhile, if people at the other end of the sensors are observing displays, we don't know how much of the context they see. We don't know much, if anything at all, about what a system is taking in about us.

On the output/actuator side, we do not always know what the system does in response to inputs. Perhaps, it changes the Heating, ventilation, and air conditioning (HVAC), or load balances electricity, or alters lighting to better see what we are doing. When humans do not know how a system responds to their actions, or they are unaware that a system is sensing and responding to them, why would they think they are interacting with something. What would be their grounds? They don't feel in control; they don't feel they are agents in dialogue, or participants in joint activity, and often they don't even know they are in the "sights" of a system and part of an interactive chain of action reaction. Arguably, the only reason to say they are interacting with the system at all is that it responds to their actions and adapts to their reactions whether they know it or not. If someone suspects they are in the presence of an adaptive or context-aware system, they may try things out. That might justify the claim of interaction. But, if the system's response relation is too complex, or the form of its response is to alter appliances they know little about, it is more of an open question whether the two are interacting. At best, the interaction would be implicit; that is, the human would implicitly know things are happening and somehow connected to their action, but this knowledge of interaction occurs without their explicit awareness. System output is too hard for us to explicitly detect it and tie it to input.

All this suggests that treating interfaces as an *n*-dimensional surface won't quite work for context aware, distributed systems, or for sensor-based AI-driven systems. There is still a distinction between agent and system, but the key features true in the classical interface of knowing the input devices, knowing their semantics, being in intentional control of what they do and when are things that no longer true for context-aware distributed systems. Because of the complex interconnections between multiple participants and processor(s), we need the expressive power of something more complex, more like a network model. See Figure 11.

See Table 2 for a comparison of the classical and network/distributed notions of an interface.

With a network model, has HCI now reached the concept of an architectural interface? Not quite. What is left out is full embodiment and the way humans interact with each other in a space they share. In particular, interaction in buildings is connected to co-presence – social, physical, epistemic, and normative. This means that in buildings, agents are an intrinsic part of their interface. They co-constitute the place they are in. In a network model, co-present agents are technically still distinct from their context-aware interfaces. They are located in one place, sensors, and outputs in another. Space is still the space of physics; it is not intrinsically a *place* [51]. Although it is not helpful to think about a context-aware interface as an n-dimensional surface, it nonetheless remains an n-dimensional interface from a purely formal viewpoint because a physical digital boundary is still present. This changes somewhat when we incorporate robotic systems into our pervasive system. Robots reach into our world more ambiguously. Depending on their form factor they begin to enter the edge of our social world. And some may have an autonomy that runs contrary to the notion of a ubiquitous system. But context-aware systems without robots still interface our commonsense world through digital input and output ports.
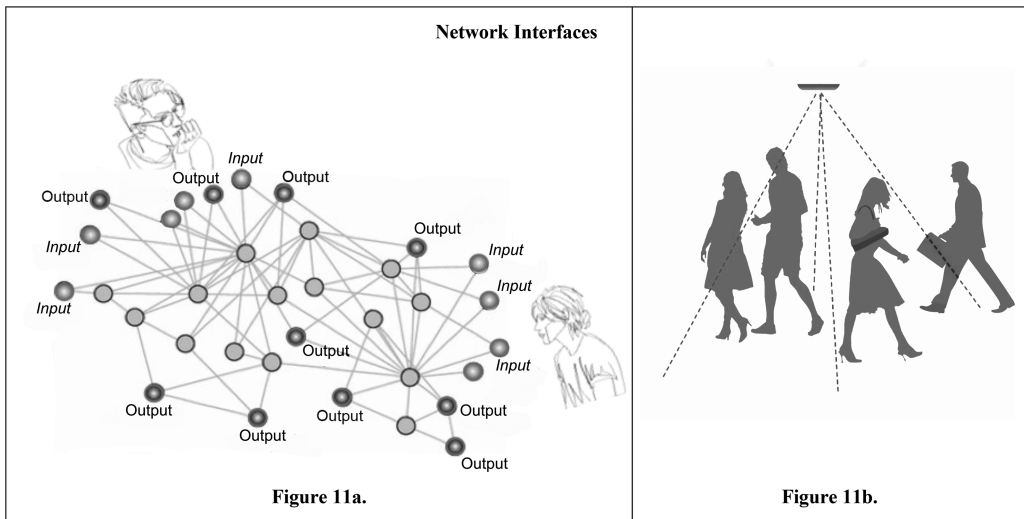
Fig. 11. (a) In a network interface, there are many lines of connectivity between input, apps, and human agents. Because human agents do not know which nodes are active at any moment there is no semantic transparency. They do not know what they are transmitting, what it might mean, what its effect might be, and they do not always know when the system is producing output. There is too much unpredictable interaction for them to really know when and how they are interacting with the system. In extreme cases, the two interact only in an implicit sense. (b) Illustrates a context-aware system tracking peoples' action in a specific region. But it may also capture their speech, body temperature, posture, and other biometric information. They probably don't know, or know precisely, what is being read off of them. If they are completely ignorant, they can't explicitly interact, and they can be said to implicitly interact only when they tacitly know things change somewhere as a consequence of walking in this area.

The bottom line is that in a network interface there is no room for the role of presence and co-presence in all its complexity. For that, we must turn to a more embodied, ecological notion consistent with architecture.

3.4   Architectural Interfaces: Embodied and Ecological

In architecture, but not in HCI (yet), people are part of their own interface. They are causally embedded in the interface, constituents of it. They have a place and presence in the interface, rather than being causes of input and recipients of output from an independent interface. This is because when two or more people are present each sees the other as one of the entities to interact with. They are players both on a stage and partly creating the stage. Following the stage metaphor, when someone opens their mobile phone or moves a chair, they alter the interface. This has weighty consequences. Since we co-create our interface, there is no full separation between interaction and interface. Whoever or whatever else is in a place – other creatures (pets), artifacts, and especially conspecifics – they are factors in the dynamics of interaction. They co-habitate with us. When they are humans, we negotiate our interface. This means we tacitly set weak limits on what we can do.

The idea that we co-construct our living space plays out in interesting ways. Constraints come in epistemic, social, physical, and cultural forms. For example, where one can sit depends on where others are sitting – a physical constraint. What one can say, read, or do depends in part on who else is in the room and our culture. How long one might have to wait depends on the queue. Most of these influences on our possible actions are not part of the network model, where agents interact

Table 2. Attributes of a Network Interface

Attribute	Direct manipulation	Network interface	What each attribute means for ubiquitous or context-aware systems
Recognizable input devices	YES	NO	Not all input devices need be recognizable. Sensors often are hidden or invisible. We don't know which are on.
Manually controlled	YES	NO	Input devices are not typically hand-sized or hand controlled – they are mostly sensors.
Feeling of control	YES	NO	You can't feel in control if you don't know the boundaries of a system: where the sensors are, what is being registered and recorded, how inputs are affected by implicit parameters like the time of day, details of the situation, whether other people are currently working with the system.
Intentional control	YES	NO	In complex systems, how can agents know what side effects their actions may have? Or what may happen given the other players, or how the system balances everything? Smart systems make decisions for us. So how can we have intentional control of output or system behavior? Often the direction of control is from system to human. Systems can initiate interaction.
Explicit understanding	YES	Sometimes	Occasionally, actions have clear meaning/results and agents explicitly know what their actions mean to the system. But actions can cascade, or lead to surprising events in the distance, or set off unintended actions by others thereby limiting predictableness and understanding. Sometimes there is explicit understanding, sometimes implicit understanding, and sometimes no understanding.
Closed	YES	NO	The entire I–O system is open on many fronts: from multiple humans, effects of timing, or random intrusions in a sensor's environment.
Circular	YES	NO	Users are not assumed to implement a function that maps system outputs to the inputs they create through acting. Too often they do not know when they are creating inputs and they have no idea what a system is doing in response. Moreover, the system is context aware and reacting to more than just that user's behavior. There is too much openness, uncertainty and user ignorance to treat users as a closed function from system output to user behavior.
Deterministic	YES	NO	From any participant's point of view, the system behaves stochastically with unpredictable events happening because of timing, collisions, load balancing, and so on. Even if all these are under the operating system's control, distant people can turn on or off their own ports and this is not predictable in advance.
Independence	YES	YES	Interactions with the interface do not change it. A system may adapt to agents by improving its service, but it does not change what inputs it registers, and what outputs it can produce, nor even where the digital–physical boundaries of input and output are located.
Direction of adaptation	Human to system	System to human	The system adapts its output to users rather than user to system. Its services are in response to user needs. Hence, advanced systems will learn users' behavior in context and adapt to these to maximize service quality.

with the network through very specific channels and the network neither co-constructs their space nor co-habits with them, though it may try to construct parts of it. The difference is clear because if someone walks out of an architectural interface it substantially changes the interface, whereas if someone walks out of range of a context-aware system the network interface remains exactly as it was. Inputs have changed but the interface remains.

One way of thinking about a more niche-like model of interface is to reflect on joint activity. Although not the standard way of thinking about joint activity, it is often constructive to think about people being together as sharing bubbles: the bubble of shared situation awareness, shared agency, shared knowledge, and common ground. Every person has social presence, epistemic
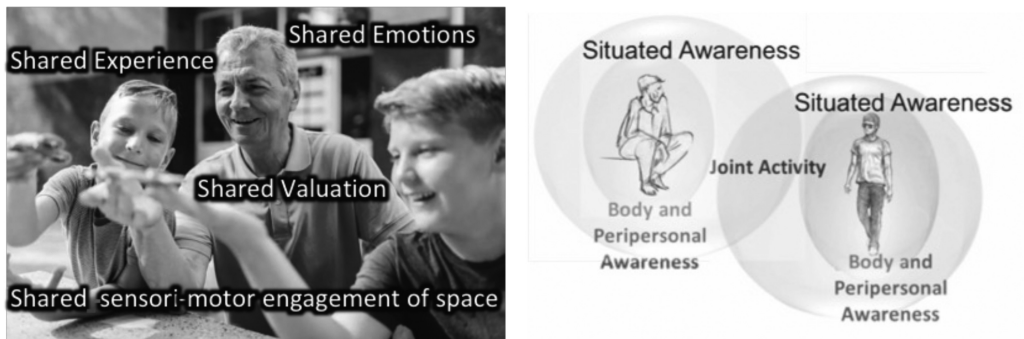
Fig. 12. In social interfaces, the participants unavoidably share bubbles of joint awareness. Each has their own body and peripersonal awareness. But each also shares knowledge of much of what the other person knows about the space they share. Because they are co-present, and being much like each other, they know roughly what the other experiences, what each is attending to, what each can be expected to do and not do, the conventions, norms, do's and don'ts that come from participation in a social group, even ad-hoc participation. Image credit: Getty free images.

presence, normative presence, and personal agency. When others are physically close, they share social co-presence, epistemic co-presence (distributed situated knowledge), and joint or distributed agency. See Figure 12. Two co-present agents share non-conceptual knowledge [16]. They each know where to look when a sound startles them. They know where the sound is, relative to themselves and relative to the other. And they know that the other knows it too. This happens without representing the sound in a Cartesian space. It is represented in each person's activity space. They also have the capacity to coordinate attention, to jointly "know where" as when both listen to the steps of an intruder [60].

Likewise, because we are human and knowledgeable, we can reason causally about events. Causal reasoning is more than statistical Bayesian reasoning [68]. That unexpected sound: what in the environment might have caused it? Knowledge of the other person: what can we infer by their implicit social interaction – their stance, aggressiveness in acting, facial gesture, position in the room, yawning [27]? Has someone entered the room? We likely know if they have, and we know that the newcomer likely knows we know it too. What can that person see? We rapidly triangulate lines of sight. We share much that is related to knowledge and situation awareness just because we are both humans and in the same space at the same time and about the same size. Context-aware systems do not share our embodied perspective. As Wittgenstein said, much of our knowledge and capacity to coordinate depends on our having a common "form of life" [85] something that robots and context aware systems do not share.

How does this play out in the attributes of an architectural interface model? First, to talk of people themselves as having input and output dimensions seems forced, inappropriate. We know where others are and a great deal about them, but that does not support knowing what and where they are as *a thing* outfitted with a set of input devices and sensors. Calling people devices, or any part of them a "device," is a category error [74]. Similar problems arise when we consider other conditions of direct manipulation such as the semantic transparency of the effects of our actions on them, our feeling of being in control [45], the two of us being a closed deterministic system. These do not apply because when we interact with other people, we do not directly manipulate them, they have freedom of will and interpretation. Nor do our fellow agents behave like context-aware systems, whose *raison d'etre* is to provide services for people. Humans, as Kant [39] famously said, are ends in themselves. Even human slaves have hidden moments of freedom when they are not

servants to others. We all are free to think as we like in our interior life and scratch, cough, choose the words we speak, and so on. Machines are not. And when we do things for people, often those things are not really services. They are things we do together; or things we implicitly do for them without realizing it. In sitting by a fire with others, we may have no intention to interact, but just by being close in space we may give others comfort or pleasure. We implicitly interact. The list of differences continues. When we act with others as a joint system, there is no obvious direction-ality of adaptation and control, as if one party adapts to the other while the other goes about her business deprived of awareness, or as if one party controls the other while the other is passive. The others are by definition 'involved'. Everything is "co": co-adaptation, co-ordinated action, shared knowledge, and co-present. Further, although there are always surprises when people interact – some degree of indeterminacy and unpredictableness – people are different than network systems because we strive for interpredictability when acting together [43]. Networks are happy to do their work without our knowledge. To date, human–human systems are essentially different than human–computer systems. Social interfaces are not HCI interfaces. See Table 3 for a list of these differences.

We turn now to the different ideas of interaction that fit these different notions of interface. How does each interface determine a set of possible interactions? What are the formal properties of interaction in each interface?

4 PART 3: INTERACTIVITY

I have been arguing that architects have a third concept of interaction, one that goes beyond the interaction of direct manipulation and networked mediation. Cognition (and brains) are different when people are nearby and share a physical space [1]. This difference provides the basis for a type of interaction that is not tied to an interface boundary and certainly does not rely on a digital/physical boundary. What is needed is a notion of interaction that works in social interfaces and reflects a non-boundary orientation.

In this part, I set out to clarify what is special about human–human co-present interaction by contrasting it with interaction as found in direct manipulation and networked mediation. I begin with a discussion of how the different conceptions of the relation "to interact with" differ with respect to symmetry, transitivity, and reflexivity, the formal attributes of any relation.

4.1 Formal Properties of Interaction

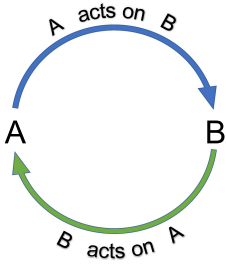*4.1.1 Symmetry.* A natural starting point for any modern theory of interactivity is with Newton and his ideas of physical interaction. In Principia [64], Newton states that interaction occurs when two bodies reciprocally act on each other. Body a exerts a force on b, and b in turn exerts an equal and opposite force on a. A acts on b, causing b's state to change, and b acts on a, causing a's state to change. Call this Newtonian condition "causal bidirectionality." Formally, it states that interaction is symmetric: whenever **A** interacts with **B**, **B** interacts with **A**. See Figure 13.

Symmetry is evident when we interact with knobs, walls, floors, or anything we touch. Touch-ing is symmetric. When a physical attribute of the thing we touch changes, our next interaction often changes. A floor when slippery creates a different interaction between floor and walker than a floor when rough. Change a key property and interaction patterns change. Stairs cause a differ-ent interaction than walls, or escalators. How a person moves in each is different because of the different shapes and textures that must be accommodated. We may expect to discover that mate-rials, shapes, tools, and furniture give rise to characteristic patterns of interaction and *a science of architecture one day ought to tell us about these patterns.*

One consequence of causal bidirectionality is that if interaction is *necessarily* bidirectional then we do *not* interact directly with many of the things we might think we do. For example, we do

Table 3. Attributes of an Architectural Interface

Attribute	Direct	Network	Architectural	What each attribute means in architectural interfaces
Recognizable input devices	YES	NO	N/A	N/A. Social interfaces can exist without digital devices. If they are present on or in people, then we assume that at least the agent wearing the device recognizes it. These devices may be part of a social interface, but they are not the focus of interaction.
Manual	YES	NO	N/A	N/A. No input devices then no manual input devices.
Feeling of control	YES	NO	Yes, but jointly with others if present	In familiar social contexts, we feel in control. But social interaction is a joint activity and hence our own control is limited by the autonomy of others in the group.
Intentional	YES	NO	Mostly	It is a convention of social life that we act intentionally on others. Implicit social interaction, such as stance, intonation, facial gesture happens all the time, however, implying that some interaction is unintentional and implicit.
Explicit understanding	YES	NO	Partial	People tend to understand their social interactions to a first order. In the context of joint activity, consequences are reasonably well defined and predictable. The track record is less impressive for people knowing the effects of their actions on others when the point of action is unconstrained by task. Actions can lead to surprising downstream events.
Closed	YES	NO	NO	The system of participants, furniture, props, and ambient "stuff" is open on many fronts: from multiple humans, effects of timing, random intrusions. Social interaction is not closed to outside and unpredictable influences.
Circular	YES	NO	NO	Users are not comprehensible as functions that map system outputs to inputs in a goal directed manner. Joint activity results in surprises.
Deterministic	YES	NO	NO	From any participant's point of view, the social interface is largely predictable because social agents when acting in a coordinated manner try to be predictable. But other participants still do unexpected things. As a system, the interface is permeable to events from outside, so it is not deterministic.
Independence	YES	YES	NO	Interactions with the interface often change it. Agents change where the interface is when they move, introduce props, change their activity.
Direction of adaptation	YES	NO	Circular or joint	Humans co-adapt.

Symmetry: A interacts with B iff B interacts with A

$$\forall a, b \; (aRb \leftrightarrow bRa)$$

Fig. 13. Symmetry in interaction means causal bidirectionality or reciprocity. This rules out one-way causation as a form or interaction, such as A being a stimulus that prompts a reaction in B. Interaction is always two-way, though the return action of B need not be equal and opposite. E.g., when a switch is turned a light may illuminate it, hence reciprocity; when a speaker chats with a signer, the reply to speech may be a sign.

not interact with TV, movies, or media when we watch or listen no matter how attentively, unless we change the volume or channel, unlike what some in media studies historically have suggested. Without our direct interference, our media continue to behave the way they would without us. Nor do we directly interact with buildings when we navigate through them, other than in the very basic interaction of walking on the floor, making sound, or creating moving shadows as we move - those some might reasonably argue that this proves bidirectionality. If accepted, this means that responding to signage, passing through open doors, and avoiding cordoned off areas is not an interaction with signage, doors, or barricades; it is a one-way response by us to cues. Action omissions, such as not hitting the door, avoiding restricted areas, going down the wrong passage, are not interactions. They are just actions. Hence, causal bidirectionality implies that we do not directly interact with many of the things in buildings that architects take pains to design. Well-designed features provide guidance rather than pushing or pulling. Good design seems to *make interaction unnecessary*.

Symmetry is a strong condition. It may exclude activities we might think are interactive.[7] On the other hand, if symmetry on its own is *sufficient* for interactivity, then it is possible we interact with many things we do not currently think we interact with. For instance, when we speak in a closed room, the room's acoustics affect our sense of how we sound. Our vocal chords affect the air which affects the walls, and when we hear the sonic rebound our internal state is affected. Speech production is affected by speech perception and that is affected by acoustics [73]. Are we interacting with the room? Most people, I expect, would require intentionality: if we intentionally test acoustics, then we do interact, but not otherwise. When speaking in order to test acoustics, we are *explicitly* listening to the response. And we get one. So yes, when we speak in a room we are interacting with the room if we intend to adapt. Others, however, might argue that intentionality is not required. We interact with a room, whether we consciously test it or not, because we

---

[7]Symmetry is an even stronger condition if we assume that interaction is non-instantaneous. Here's why. If A → B takes time, then B changes state after A started the interaction. That means that B → A will change A later. Since symmetry means that it doesn't matter whether we start with B → A or A → B, isn't there a danger that we are committed to interaction (or causation) being an unending loop of A interacts with B which means that B interacts with A, which means that A interacts with B and not just logically but physically, hence leading A and B to loop interactively forever – like gravity? This raises the natural question whether interaction implies that two things must have more than one-shot bidirectional causation. That's a worthy question since we tend to assume that if A interacts with B then A also responds to B's return action on A and in a manner that leads to B's renewed reaction. But we need not let a concern with eternal interaction derail us if we focus on a logical analysis of the meaning of the relation in abstraction from time. If interaction is a relation, it can be both symmetric and have a starting direction, only were it a function or operator would it be commutative, implying that no side has causal priority.

implicitly react to acoustics by unconsciously modifying how we speak. In that case, there is a loop of causation: from us to the room, from the room back to us, and then through our adaptation of speech back to the room. It can be argued that this is sufficient for interaction because if we consider a contrived situation where we cannot hear ourselves speak our natural flow may be partially derailed, proving that we rely on adapting to acoustics to implicitly control our volume and clarity. Realize it or not, we implicitly interact with rooms when we speak.

These two consequences of causal bidirectionality – that necessity rules out some cases we might intuitively include and that sufficiency rules in some cases we might intuitively exclude – highlight the complex role that intention, awareness, and sensitivity play in interaction, especially in implicit interaction. In HCI, the classic account assumes agency because it is inspired by direct manipulation: a user intentionally acts on an input device to create changes in a target system [36]. This presupposes that users also have the sensitivity to be aware of the effects of their acting. It is a good start and works well for knobs and elevators, drawers and bathrooms. But it works less well for many architectural elements that architects care about because no one thinks of things like materials, textures, colors, openings, and shapes as input devices. They are indeed features of the interface, but they are not devices that we act on or try to change. We are typically passive to them. We register them and respond to them (often unconsciously) rather than try to change them explicitly. These adaptations, as with our speech example, might mean we do interact with architectural elements, albeit implicitly, despite our being unaware that we act on them, and unaware of the effects our actions have on them. If we do interact with them, it is likely that the nature of our interaction is subtler, less obvious, and might well require scientific study to discover and understand. The same concerns about causal bidirectionality and explicit agency arise for pervasive computing where contextual sensors are ubiquitous and for everyday human–social interaction, where a network notion or something ecological may be required rather than an explicit "I push on it and it pushes back" notion. Explicit agency may be overrated for interaction.

One reason explicit agency is indeed overrated, I believe, is that it is easy to underappreciate how widespread implicit agency is in our social life where we both transmit and receive implicit (unintentional) social cues and social attitudes all the time. These cues are displayed in posture, movement dynamics, interruptions, pauses, prosody, and joint gaze. Members of a co-present group do not have to be aware of the cues they each are transmitting for those cues to have social significance and affect how each member reacts [71]. The process recurses, with reactions feeding and amplifying other reactions. In the theory of social interaction, such unconscious, unintentional cueing is an active area of research [23, 86]. It is clear that in the social world we implicitly interact.

Why should implicit interaction be acceptable for social interaction but unacceptable for other forms of interaction? Isn't it more reasonable to assume that neither intention nor awareness nor a sense of agency is necessary for interaction as long as all parties involved in the interaction are relevantly affected and, if human, then implicitly "aware"?

Where does this leave us? Bidirectionality has given us implicit interaction. Good. But it has rejected most building navigation as a form of interaction. Possibly bad. At least half of my architecture informants and about half of architecture grad classes when polled (about 75 students), all believed that navigating through a building is a major way of interacting with it. Assuming those architects are not just "the half that got it wrong" is there any way of saving navigation while retaining causal bidirectionality? To put this in concrete terms, can we explain how it makes sense to say that inhabitants of a building interact with walls and doors when they seamlessly move through, never touching or bumping them?

One possible resolution is to see the problem arising from a simplified notion of causation. Move beyond that and we can rethink casual bidirectionality, network interaction, and human

co-present interaction. The first step in that direction is to clarify transitivity and explain how mediated interaction works.

4.1.2   Transitivity. Interaction is transitive when we control things through the use of other things and receive feedback or a return action in some form. Transitivity means that when a interacts with b, and b interacts in the relevant way with c, then a interacts with c. $\forall a, b, c (aRb \land bRc) \Rightarrow aRc$, given appropriate conditions.[8] Interaction can be mediated. It often is. For example, when a tennis player makes an enviable shot, placing the ball in the cross-court corner, she may justifiably be said to have interacted with the ball. But not directly. The racket is a mediating instrument through which the player interacts with the ball. When playing is fluent, the racket is transparent to the player since her or his entire concentration is on what to do with the ball: where to place it, what sort of spin or bounce to make it have. Causation and control *pass through* the racket enroute to the ball.

The phrase "appropriate conditions" in the transitivity condition plays a vital role. Many actions that support interaction are transitive only under certain conditions. If a walks with b and b walks with c, then a walks with c, assuming their walking is simultaneous. Walking-with is a joint activity, and it is symmetric[9] like playing tennis together or participating in a tug of war. Does the joint activity of walking-with apply to a hundred people or to people so distributed that there is no control, coordination, or communication between them? At some point, the transitivity of walking-with breaks down. The conditions of application must be defined.

Consider these problems:

- *Long causal chaining.* Can a cause something that ripples transitively from $b \rightarrow c \rightarrow d \rightarrow \dots \rightarrow z$, such as a's veering sideways or increasing pace? If a's impact on b washes out before reaching z, how can a possibly interact with z? What determines the length of a causal chain? If a cannot have an effect on z, how can a and z interact?
- *Epistemic requirements of forming an intention.* Suppose a has a causal effect on z but doesn't know it. Can a interact with z if a does not know of the existence of z? We're assuming that a doesn't need to know much. It would be enough that a is able to refer to z as "the person on the extreme right," or "the person just after the farthest one I can see"? But if a does not know that z is present, or if a lacks the referential capacity to think about z as an individual thing or person, how could a form an intention to interact with z? Even implicit intentions require the capacity to know (albeit implicitly) the difference between affecting z and not. At a minimum, a must be able to detect feedback from z. Is there some symmetric ripple back effect that a could pick up, even implicitly, that would serve to close the loop on interaction? It can be something as small as a complaint from z, or a groan, or a stumble that somehow in some form gets back to a and it carries enough signal for a to know that s/he was in causal connection with z (wherever z might be).

---

[8]Although causation lies near the heart of interaction, and causation is usually thought to be non-transitive, this is one respect in which interaction is not *just* bidirectional causation [Halpern & Pearl, 2005; Halpern, 2016a]. See Lewis (1973) for a defense of causal transitivity.
[9]All interactivity is symmetric, but it is symmetric in the sense that b's return action must change a's state though not necessarily in the way a changes b's state. When two people walk together, a walks with b iff b walks with a. They symmetrically act in the narrowest sense of both acting on the other through walking with. If a directs b, then a interacts with b but b's action back on a may be that b's changed activity is visible to a. This satisfies the condition that b provides a with feedback. Feedback may be perceptual or non-perceptual, e.g., changes a's non-perceptual physiological state. *Walking with* also provides feedback in that b must walk within certain bounds of a and a must walk within certain bounds of b, so the two dynamically entrain each other. Typically, this will involve perception but entrainment per se does not require perception. So the feedback from b might be in terms of force, pressure, holding hands, and this might support long-term interaction past the point of perception.

The same question arises for every action we might consider as being interactive. If *a* is talking to *b* and *b* to *c*, can we infer that *a* is talking to *c* through *b*, or that *a, b, c* form a group where each speaks to every other? It depends. Talking is also a joint activity, but the rules of who, how, and when people can "talk to" each other are different than the rules of who, how, and when people can "walk with" each other. We need constraints on when talking to is transitive. These too are part of the conditions of application we will have to define.

What about purely mechanical actions like steering? To control a car, a driver must interact with the steering system. The steering wheel is connected to the steering system, usually through pneumatics, linkages, and gears eventually to the front tire angle. As the car turns, the driver receives feedback. Changes in the car's angular velocity cause the feeling of acceleration; force feedback from the steering wheel often gives a sense of wheel resistance, its arc gives a sense of turning sharpness, and the movement as the car itself turns changes optic flow. This raises an odd question. Because of the chain of internal parts, all playing a role in turning the tires it follows that a driver cannot turn a car without affecting all the intermediate links in the chain. Does the driver interact with those internal parts explicitly, implicitly, or not at all? The answer, predictably, depends on the nature of feedback or return action from those intermediate parts. Defining the requirements for chaining of mechanical interaction is another instance of the need for a deeper analysis of the conditions when transitivity holds.

Our analysis of transitivity is becoming complex because interaction has several components which until now we have only incidentally distinguished. To be more precise, if *a* explicitly interacts with *z*, regardless of whether *z* is nearby or is a distant thing and *a* interacts with *z* through intermediaries *b, c, … y*, then *a* must have the following:

(1) *The capacity to refer to z. a* must be able to form a thought about *z, a* must be able to form an intention to interact with *z,* and this epistemic capacity implies that *a* has concepts or tacit categories that fit *z* in some way.

(2) *The capacity to discern feedback from z. a* must also be able to pick up information from *z,* so *a* can distinguish doing things that have an effect on *z* from doing things that have no effect on *z*. This is a second, though related, epistemic capacity that enables *a* to correctly identify feedback as feedback *from z*.

(3) *Causal power over z. a*'s actions must actually have a causal effect on *z*. This is an objective fact about *a*'s control over things like *z*.

Applying these conditions or capacities to the car driver: if we learn *a* is ignorant of the existence of deeply internal steering parts, regardless of how significant their activity is for proper control, how can *a* have thoughts about those parts, and so have explicit agency? To those who are not car mechanics, the inner parts of a steering assembly are conceptually invisible. They are just steering stuff. Without any notion of what the stuff is, how can someone set out to interact with a specific part of the stuff? Explicit agency requires explicit intentions. And explicit intentions require having an explicit object of thought.

What about implicit interaction: might a driver interact implicitly with inner parts of the steering system even without knowing the parts? The implicit knowledge part of interaction mostly concerns making sense of feedback. A person who knows little about cars may still unconsciously pick up cues about moving parts inside the steering system. Perhaps, the car pulls to one side in specific circumstances. Or a funny sound seems to come from the wheel area when turning suddenly. These cues may be tuned into by a person's implicit knowledge systems [82] long before they are grasped as something identifiable by the explicit system. They can have an effect on intentionality too because without realizing it a person may adapt their driving to prevent those sounds.

So someone may have an implicit idea of "the thing making that odd sound" and without thinking about it explicitly they drive to minimize the sound. Condition one can be satisfied. Such drivers also are sensitive to feedback because they can implicitly discern feedback from "the thing that makes that odd sound." Condition two can be satisfied. The final condition of causal efficacy can also be met because by definition turning the steering wheel affects internal parts. Causal power is an objective fact of causal connectivity, and again by the way we have posed the problem, they are causally connected. Accordingly, a driver might under these conditions interact implicitly with internal parts without realizing it. By satisfying all three conditions, a driver interacts with hidden parts of the steering assembly. As to whether the interaction is explicit or implicit: the answer will depend on whether the person's beliefs about noisy internal parts are explicit or implicit – if implicit, then they are things the driver knows about his car without realizing that he knows it.

These conditions sharpen the notion of interaction as happening, at times, through mediation. But there is more to be said about mediation. We still have not explained the difference between a person who interacts with z and feels s/he is interacting *directly* with it and a person who feels s/he is interacting with z but only *mediately*, through a set of intermediaries. Both are cases of mediated interaction but one feels like it whereas the other does not. The idea of control transparency discussed earlier requires feeling as if one is acting directly on a third thing, *c.* There is a chain of intermediaries, but the chain is phenomenologically transparent. We see right through the chain, and act right through it, with the sense that we are acting on the end target, z, directly, much the way we see and act through a glove when picking up a board.

4.1.2.1   Transitivity and Transparency. Control transparency is the counterpart to perceptual transparency – seeing through – that Merleau-Ponty [57] discussed with his example of long canes. The two are connected because one never just sees through an intermediary – seeing right past it to some more distant perceptual object – without also transparently controlling the intermediary. And one never transparently controls through an intermediary without also transparently sensing feedback from it. Control transparency implies transparency of seeing and seeing through implies transparently controlling through.

Take the case of binoculars. When we look through binoculars, after a brief moment of adaptation, we become a binocular control system: moving our hands to ensure the binoculars scan or track items far beyond what we can see with unaided eye. We don't interpret our hand movements in measures appropriate to the space around us, where things are understood in dimensional units geared to hands and arms, the actuators intrinsic to peripersonal movement. Our *sensori-motor control* of binoculars is tuned to the features of the target world, the remote world. Thus, we don't think about inching the binoculars to the right or left in local inches. We think about the target domain; moving a few inches along the branch or a meter over to the next tree while watching the movement of a bird. Any explicit intentions we have are rooted in concepts that are meaningful in the target domain. If the binoculars could somehow magnify like a microscope, then our distant domain would be the tiny world of bacteria and textures specific to that world and our interpretation of lens adjustments and focus would be in terms of clarity in the microscopic domain. Only if things go wrong, do we momentarily shift attention from the remote to the local world [28].

Binoculars and long sticks are extensions of our sensory system. We interact with them, through them, to perceive, to extend sight. But they were not designed to let us cause change in remote things. Binoculars, hearing aids and microscopes leave their targets unchanged. Hence we do not interact with the birds we see when we track them with binoculars. Binoculars show that when we see through something we inevitably transparently control it. But they do not show the opposite side of the equation: when we control through something we also see through it.
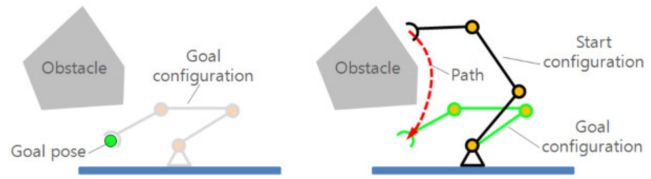
Fig 14a: Seeing through   **Fig 14b: Controlling through**

Fig. 14. (a) We *see through* a laser pointer because when pointing, we don't think about our hand, just the remote surface. In robotics, there are two problems: planning the position and orientation of the end-effector in Cartesian space (the pinpoint of light on a surface) and solving the inverse kinematics problem of computing from the Cartesian path the light creates on a shaped surface to the corresponding trajectory of the manipulator in the configuration space. (b) A robot arm with four joints in its start and goal configuration. The obstacle constrains the viable paths. Once a path is chosen in Cartesian space, the path must be translated into a path in the space of multidimensional joint angles. Forces too must be computed at joints. In humans, the analogy is with thinking about the path the laser will mark out and letting our motor cortex determine how to move the laser pointer to achieve that effect. Image credit: http://www.coppeliarobotics.com.

To explain this, we must go beyond the standard discussion of the seeing through type of transparency. When do we feel we can act at a distance with the same control that we have with things that are at hand? Transparent control must feel intuitive.

Let us define a *system coupling* as the conditions that ensure there is a tight enough causal connection between us and the target that we can both see through and *directly* control outcomes through intermediate links in a transitive chain of interaction. We can interpret our actions in terms of their distal effects on the remote target. The feedback we receive about our actions is reliable and complete enough that it is comparable to acting on local objects. This means that we know how to change the remote thing without thinking, and we understand those remote changes as the natural outcome of our actions. Thus, when we turn the TV off by hitting 'off' on our remote, we understand our local action in terms of its remote effects. action despite the extensive electronic mediation.

A formal account of this sort of 'controlling through' algorithm is found in robotics where there is distinction between thinking in the "goal" space and thinking in the "configuration" space [47]. In using a laser pointer to trace a contour on a screen, my thoughts are on deciding where to shine the light, not how to move the physical pointer, using my shoulder, arm, elbow, wrist, and fingers. I don't give a thought to my motor system or how it solves the inverse kinematics problem of translating my chosen pointer path in the external goal space to internal joint angles and muscle forces [40]. Unless there were an appropriate system coupling between pointer path and my muscular control of the pointer's angle over time, we couldn't have this transparency. There must be a mapping from joint configurations to goal space that supports the desired movements in goal space. And the computations necessary to solve the inverse kinematics problem must be reliable and unconscious else we will shift our awareness from the end point of the light to our joint positions and movements. Only when things just seem to take care of themselves do I have control transparency [4]. See Figure 14. Like our binocular example, our sensori-motor sense of where we are pointing and how far things are from each other now presupposes the inverse-kinematics problem is solved, both for planning, executing, and interpreting. It explains why we can see through our

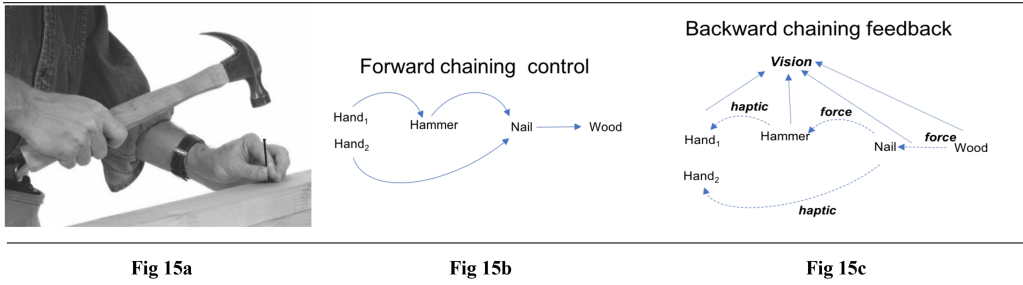| Fig 15a | Fig 15b | Fig 15c |

Fig. 15 (a) Perceiving through hammering is more complex than perceiving through a white cane because all parts must be appropriately coordinated – made into a human-hammering control system. *Controlling through* is a new term to describe the capacity and accompanying feeling that people have when they feel in charge of a process because of their skill in creating a reliable set of couplings. They feel they are tightly coupled with each link in a causal chain – hammer, nail, wood – hence, they feel they are responsible for the process. They are the controlling part of a causal system. (b) We see the forward side of hammering. One hand holds the hammer, the other the nail against the wood. When things are under control causal forces move forward as they should. (c) We see the feedback side of hammering that is also necessary for control. Vision enables us to directly see the hammer striking the nail, while haptic sensing via our nail holding hand (hand$_2$) enables us to directly feel the nail's position on the wood. Haptic sensing also lets us discern distal forces as these propagate back from wood to nail to hammer. As our skill in hammering increases, we eventually sense and control things well enough to become a transparent hammering control system. It feels to us as if we see and control through the hammer so effortlessly and so successfully that we *transparently* act on the nail – we act at a distance.

joints and laser pointer to the dot on the screen, and why we can transparently control that dot. The only difference between binoculars and laser pointer is that with a laser pointer we have a remote effect on the thing we are monitoring.

Let's apply this seeing and controlling through analysis to the transitivity of hammering a nail. First, I must have in mind how I want the nail to move through the *wood*, the angle, the speed, what to watch out for – these are my seeing through concerns. They only bear on the goal space: hammering that nail in correctly. My attention soon shifts, however, to my grip and how to swing the hammer to impact the nail. I can't *transparently* control through the hammer yet because I have concerns about controlling through the hammer that must be resolved before I feel I can act at a distance – before I can transparently control the nail through the hammer. Because a nail can shift position, I also must hold the nail rigidly to preserve the system coupling between nail, hammer, and wood. That means that to feel in control of the coupling, there is a chain of additional interactions beyond hand gripping hammer that I need to regulate, specifically, ensuring the hammer strikes the nail correctly and the nail remains correctly aligned with the wood. See Figure 15. Controlling through refers to the extra control I must exert to ensure the causal chain works right. It depends on maintaining the right causal coupling and registering feedback from links in the chain. I have to create and maintain a system coupling.

This highlights two points: (a) transitivity presupposes a reliable system coupling; (b) sometimes this happens automatically and becomes transparent without our conscious involvement or concern, as is the case with the joints in our arm, the modern laser mouse, the steering system in a car. At other times, we must consciously create and maintain a reliable system coupling between multiple parts, as is the case with hammer and nail, where we need to control three things: our grasp, the hammer-nail dynamics, and the nail-wood dynamics. When everything goes well, and we execute the coupling well, we may begin to attain control transparency and see through

hammer and nail to track the nail's progress in wood. Of course, our eyes help here. Our visual monitoring of the nail in wood is independent of our tactile sensori-motor control of hammering. Mediated interactivity may depend on many modalities. The moral, though, is that creating and maintaining a system coupling is central to reliably controlling through. When it becomes automatized, it enables control transparency. But it may never become automatized and we may never get to the point where we make unconscious adjustments to keep things on track. Well-designed tools, as most hardware tools are by now, provide the necessary affordances for achieving a measure of control transparency. But as the number of links in a control process increases, we may need to bring in helper tools to create a system coupling. This system coupling is the causal matrix that guarantees chained feedback and a reliable form of chained forward force. In such cases, a person who knows how to manage the links in the coupling will feel in charge of the coupling; they will have a sense of controlling-through the couplings, producing an alignment between intention and execution. They have explicit agency and explicit transitive interactivity. In the best cases, these system couplings, despite their complexity, become transparent.

4.1.3 Reflexivity. Reflexivity is the last formal attribute to discuss before we can apply our analysis of interactivity to HCI and architecture. When a system can interact with itself, such as when we scratch ourselves, talk to ourselves, warm ourselves up by running, we interact with ourselves reflexively: $\forall x: xRx$. Reflexivity is a form of unmediated action on oneself. Walking is not a reflexive type of interaction because we cannot walk on ourselves or with ourselves. Most models of interactivity do not support reflexivity because causation is not reflexive – things are rarely self-causing.

It is easy to think that reflexivity is nothing special. Doesn't a system interact with itself by relying on symmetry and transitivity? For instance, dancers regularly seem to dance with themselves when in front of a mirror. But not the way two physical partners would. Their dancing with themselves is mediated by their image: They act on the mirror, the mirror acts on them through their perception, and they react to what they see. Owing to transitivity they act on themselves mediately. So dancing with oneself via a mirror cannot be reflexive interaction. It is still a form of interacting with oneself; it is just not reflexively dancing with oneself.

Dancing with another person is different. When two people dance together each is dancing with the other, so they can change the system consisting of the two dancing by changing their own movement. An individual dancer's modification (stepping left instead of right) has a global effect on joint dancing. Especially since the other is likely to adapt. When they change themselves, they change the system they are part of.

Few systems support genuine reflexive interaction. In particular, neither classical nor networked HCI do because both those types of interface distinguish the agent from the system. There is still a distinction, a boundary, between where input enters a networked system and the agent creating the input. People cannot be a component part of a network interface; they are on the outside. It is the network that determines where the interface is. Anything people do with themselves, say scratch or mutter, may be picked up by sensors as inputs to the network system. But that is not good enough for reflexivity. Interaction with the system via its digital sensors is mediated; it is no different than acting on any other input device. True reflexivity, by contrast, requires that the very act of interacting with oneself is an act of interacting with the system without mediation. Only when a person is a *constituent* of the interface – a part of the overall system – can they change the system without mediation. That is why scratching oneself is reflexive. An arm is part of a body, so one part's action affects the whole without mediation. It is also why playing tug of war is reflexive, since one person's extra tug is a change to the whole system as well as to themselves, or why two people talking is reflexive because any one person's increase in volume is an increase in the volume
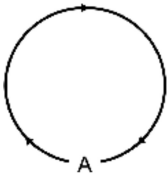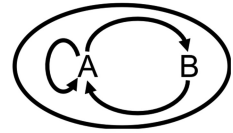
| **Fig 16a** | **Fig 16b** | **Figure 16c** | **Fig 16d** |

Fig. 16. (a) Reflexive interaction occurs when an entity acts on itself, thereby changing its own state. It typically requires that a part of a system act on the system as a whole. (b) Scratching oneself is reflexive interaction because one acts on oneself without the help of tools, such as a comb. Hands are part of the bigger system – the body. Scratching is both a change in a body part (it is moving) and a change in the body – scratching gives relief. (c) As one dancer changes her steps or posture, she changes her relations to her partner and their joint position. In social interaction, individual people are part of a bigger system – a social group. Some of their individual actions constitute social interactions with the group, possibly changing it. (d) schematically represents how dancer A when moving to the right, or when putting on new shoes, effectively acts on his own state; this has an effect on dancer B whose response further affects A. The complete set of actions and reactions shapes their duet, as represented by the outer ellipse.

of the two, or one person's quickened speech is a change in the system's communication rate. See Figure 16.

With these properties of interaction in mind, we can now characterize the type of interaction that goes with each interface type. Since buildings involve all three types of interfaces, and all three types of interaction we now will have a more complete account of how humans interact with buildings.

4.2 Direct Manipulation Interaction

In architecture, direct manipulation correctly characterizes our everyday causal interaction with manipulable elements – switches, furniture, plumbing, and doors. We have other ways of interacting with buildings and building parts, but when facing the purely physical parts of a building in one on one interaction, we primarily interact directly.

Direct manipulation is symmetric and transitive, but not reflexive. The symmetry of our physical interaction is obvious, Newtonian bidirectionality. When we act on our own, without anyone nearby, interaction is essentially push me pull you. Transitivity is obvious too because so much of what we do involves causal *systems*. Turning a knob turns the oven on, it begins heating the interior and starts to cook the casserole. It is a system. To make that chaining of process transparent, our devices behave predictably. Well-designed direct manipulation systems have predictable system couplings; they maintain a lawful connection between input actions, system state, and appropriate feedback. No surprises here. Ovens, voice-controlled TV's, light systems, and so on, all fit the direct manipulation model. Do this here, that happens there. Feedback happens through the immediate visibility of effects, or through delayed but reliable response (the oven heats up), through sound, or through onscreen displays.

Reflexive interaction, however, is not supported in direct manipulation systems. Physical interaction could, in principle, be reflexive if it is part of a bigger system of interaction. Sometimes it is. But typically, physical interaction occurs at an interface where an agent *a* acts on physical thing *b* and *b* in turn acts on *a*. The two interactants are distinct. One is a human capable of self-caused actions, the other a physical thing. Both have separate trajectories. Thus, interaction between person and thing either happens at an interface – the knobs, floor, seats, or windows; or it happens through mediation as is the case with digital systems, where we interact with our TV's, HVAC's,

wireless lights through input devices that mediate our control. The only way we might interact with physical things in a reflexive manner is if we act on them without an interface. For that to happen the tool, object, or artifact must be absorbed into the neural representation of the agent's body schema [52]. At that point, it becomes inappropriate to speak of direct manipulation because to paraphrase Merleau-Ponty [57] how we control our hands is not how we control things outside us. Body parts are not in space in the same way as things that are not part of us are part of space.

4.3 Network Interaction

In network interaction, multiple components act and sense non-sequentially. It is hard to know when and where input is being gathered and when and where output is being returned. As more sensors are being incorporated into buildings, the more our interaction with buildings becomes network interaction.

A familiar complaint about sensor-based and context-aware systems is that humans do not feel in control [26]. How can someone feel in control if they don't know where sensors are, whether they are on, or what they are capturing? Is their gesture being interpreted by cameras? Are inferences made from their speech intonation? More often than not input is initiated by the system; people are frequently the patient not the agent. On the output side, things are no more intelligible. A network may respond to an input in diverse ways. It might do any one of A or B or C, or all of them. Disjunctive output is hard to notice, and disjunctive concepts are intrinsically more difficult to learn [21]. This makes recognizing how a system responds to one's action – on those occasions when one knows that one's actions are being captured by a sensor – hard to detect and understand. Indeed, the response function of a context-aware system is usually unlearnable without large learning sets, which agents may never encounter. Essentially, occupants of a context-aware system are either ignorant of the system's actions or they feel like they are flying blind. They regularly have no idea how to make the system serve their needs.

Why say that person and system interact if the person is unaware of their effect on a system, and unaware of its response? People need to know how to introduce inputs and cause change if they are to sensibly form explicit intentions. Without explicit intention there can be no explicit agency. The simplest notion of human non-human interaction is when the human acts intentionally as a voluntary agent. Yet in the social world, interaction does not have to be voluntary, intentional, or explicit; people can interact implicitly. When people are together, whether or not they are aware of their own implicit action, or of the other's, or of the bidirectional relation holding between their joint implicit actions, it is still correct to say the two are interacting. They implicitly interact when they share space together because they mutually affect each other in a recursive manner. Think how meaning laden body posture, word choice, body noises, and social distance is. People can interact without explicitly knowing they are.

Why should it be different when the thing acted on is an invisible context-aware system? This is the question that arises for network interaction. The obvious reason not to reserve implicit interaction for human–human interaction alone is that bidirectional causation – symmetry – is sufficient for interactivity. But that reason is not quite adequate. A caveat is needed: the return action of the affected system must in principle be discernable and intelligible. The system or thing responding to an agent's action – the interactant – must produce a learnable response. This extra condition is needed because symmetry on its own says nothing about the detectability and comprehensibility of return actions. Bidirectional causation requires only that there be a causal law governing how the interactant reacts and that this reaction must reach the original agent. The law and reaction might be arbitrarily complex or arbitrarily small. In extreme cases, the return action might be undetectable or so chaotic that it is humanly unlearnable. But then how could a human know that something has occurred? The main idea of implicit interaction is that explicit knowledge of

bidirectional causation is not necessary for bidirectionality to occur. Knowledge of bidirectionality may be implicit. This still requires that the function describing the interactant's response be statistically learnable. It may be hard to learn and elude learning through explicit strategies. But if agents manage to learn it implicitly then they can respond adaptively and so control or partly control the interactant unconsciously. This means that a person doesn't have to have explicitly intentional control to have control – *implicit control* [58]. Accordingly, bidirectional causation is sufficient as long as the function describing that causation is learnable. Interaction is an objective fact of reciprocal responsiveness, not of awareness or explicit belief.

If this analysis is on the right track, it means that people can interact in context-aware systems even when the system is pervasive and seemingly invisible. It is irrelevant how many steps a network transitions through or how many steps there are in its transitive chaining of b causes c causes d. As long as the overall function that maps inputs to the eventual outputs that reach people is reliably implemented and is implicitly learnable people can still interact with it despite being unaware that they are able to. Agents can interact implicitly.

Here is an example to support this view. Imagine a case where an agent is unable to explicitly detect whether some of his actions cause system outputs. Perhaps, those outputs are lagged or disjunctive or come in a form that is unexpected. The network system reliably produces a response; it just goes unnoticed. The person, in fact, is impacted. But he doesn't register that impact and so does not see it as a response to anything he did. For example, in some buildings, the last person to leave and the first person to arrive trigger an auto-off/auto-on switch in the HVAC. Suppose that last person is me, how would I know I caused the heating or AC to go off? I wouldn't. I'd know when I closed the door; but why think that triggered anything? If I am the early bird, the lag between the HVAC turning on and my zone cooling or heating up would likely prevent me from confronting my effect on the system. Imagine now that the HVAC does make a detectable sound; but either I fail to notice it or fail to connect that forgettable sound with the HVAC or with something I did. If this happens regularly, my implicit system may pick up a regularity and implicitly learn there is a connection between my early arrival and the HVAC. In that case, I've interacted with the building implicitly. I acted on the building and it symmetrically acted on me. Although I had no idea that I was affecting the HVAC I did implicitly detect it. According to current theories of explicit vs. implicit agency, having an active concept of what one wants to control is necessary for explicit agency [79]. To form an explicit intention to do something requires explicitly knowing what you want to do and achieve. Here, we have no such explicit judgment about cause and effect. But we do have correlational knowledge. This is sufficient to form implicit intentions and implicit agency, however, because making an implicit connection does not require explicit, articulable understanding.

Let's take the case further. Now I arrive early in the morning, and this time, I do in fact notice a sound and notice that the heat soon comes on as soon as I open and close the door. On a hunch I check in the evening. I hear a click, and then I hear the fans powering down. Suddenly, what was tacit is no longer. I have a conjecture about the effects of my action, and I have confirmed it. My explicit knowledge has changed. I now intentionally interact ever after. Of course, the feedback is no more observable than it was before. It was there the whole time. The difference is that now I notice it. This shows that we need to dissociate interaction and explicit agency. Else explicit knowledge alone would change the truth about interaction while leaving control, and adaptive response to feedback intact.

Following this thought, let us define implicit interaction as interaction that does not require explicit agency. It still requires implicit recognition of an association between action and some sort of feedback. What extra does agency – explicit control – confer? In a word, planning – the apparatus for deliberate action and intervention that comes with conceptualization and reasoning

about effects. Once I know what my actions do, I know how to cause the lights to turn on and off. Had my interaction been implicit I could not have made plans involving turning the lights on or off via entry. My capacity to do things with the system increases as soon as I move from implicit to explicit interaction.

Being in a context-aware system is not the only time we should think about interfaces and interactivity as a network and think about implicit interactivity. Arguably, network interactivity defines much of the everyday context of inhabiting a building where we interact in both explicit and implicit ways.

Imagine your own living room. Surrounding you is furniture, walls, functional artifacts, and various things you might use for entertainment or as props during social interaction. Direct manipulation nicely covers your manual interactions, but what about interaction at a larger scale? For instance, suppose you move furniture around. Which type of interaction is this: direct or network?

It depends. Often when a person shifts a chair or side table is their real concern with the arrangement. Arrangements usually include more pieces than the few one might touch. It is a property of a collection – a global property. This invites a question. When you touch one or two pieces are you interacting with the whole configuration – the collection – or just the pieces you actually touch?

One reason to think you are interacting with the whole configuration is that in musical performance one interacts with the sounds coming from the group – all the other instruments and not solely with one's own instrument. Jointly, the group creates chords, harmonies, or syncopation when the sounds combine. Emergent musical structures would not exist without the contribution of each performer. Moreover, the interaction is social and involves distributed cognition [44]; players co-adapt to each other's style and activity. Even when practicing alone in a practice room with a recording of others playing – a situation more like moving furniture because canned music is not responsive – the sound still blends with the others, suggesting that I interact with more than just my own instrument. Perceptually, both live and music-minus-one[10] situations are broadly similar. The walls and other items in both rooms are acoustically affected in a similar way, supporting the idea that I am interacting in an objective sense with the whole system of walls, space, sound, and my instrument. The interaction with other sounds is not just in my brain; it is an objective property.

Might the same also be said about moving furniture? I touch one or two pieces, but I decide where I have to move the pieces and how to adjust them on the basis of the whole configuration. By acting on one piece, I affect its relations to all, and so transitively one might argue I interact with the group too. They constitute a system. In actor network theory, non-digital artifacts can combine to form artifact ecologies, where changes in one are thought to propagate to others [46]. The internet of things (IOT) with its constant message passing makes that position more credible, but it may still apply without the IOT in many physical contexts.

Do network models of interaction offer an answer to the architectural dispute about whether occupants interact with walls when they navigate? The question is more than cocktail amusement. Architects passionately believe that when they design the shape of a room – say, making it round vs. square – they affect the interaction of inhabitants. This has always puzzled me. Do they mean (a) the shape affects interaction at certain times, while at other times it does not; or (b) the shape affects interaction every moment but in a probabilistic way? There are possible arguments for each side.

The obvious version of position (a) is trivial. We know that architects keep wall collision in mind because they design walls without protrusions to prevent injury. Whenever a person walks from

---

[10]Music-minus-one is the name given to recordings of concerti and other classical pieces where the part of the soloist or other key player is not recorded and musicians at home have the chance of feeling they are the performer with orchestral or ensemble backup.

Figure 17a   Figure 17b

Fig. 17. (a) Are the people in this art gallery interacting with the walls when they move through the space without bumping into walls? (b) Are the people in the center of this crush of people now interacting with the wall when they shift about in contact with each other and in partial response to how the outer folk bump into and move along the walls? Images: 17a Fotolia_Dmitry Vereshchagin; 17b Geoff Robinson.

one point to another in a building, there is a small but real chance of scuffing or bumping into a wall. When there is a collision, an interaction occurs that architects want their design to avoid. Wall shape affects interaction at certain times, at other times it does not.

Yet, if this were all there is to it why should an architect be proud of a wall design that does nothing more than minimizing occupants' chance of collision? Surely, their pride and belief in a wall's interactivity during navigation, is based on something stronger, such as the way the wall design affects the occupants' social behavior or state of mind. This more reasonable ground for pride suggests that walls can have additional effects on us beyond physical impact. They may shape inhabitants' navigational experience – the pleasure of following a slow curve or discovering an elegant shortcut. Or perhaps, the interaction is socially mediated? I choose a corner of the room to stand in and talk with a friend. Neither of us touches the wall though we come close. Clearly, the wall acts on us because, for example, we always sense it and so it partly shapes how we interact with each other, how we position ourselves, how we move relative to it and others. Our response may not be called navigational, but it is certainly in the architectural spirit of interaction. But the question remains: how do *we* affect the wall?

Here is where our position might get subtle. We interact with the wall cumulatively, a tiny bit each time, but unnoticeably. If we move one way because the wall is round and another because it is rectangular, we no doubt cause different wear patterns on the floor, we dirty it in different ways, and in the long run, we may force a change in the lighting or sound or some other way we might affect the wall. These noticeable effects are not one-off, but cumulative and lagged. Why should interaction with insensate objects require that bidirectional causation be large enough for the human agent to detect it *each time*?

Position b is more compelling. We affect the wall in a more substantial way each time but probabilistically. The argument now is that we always interact with the wall because there is always a chance that we might touch it. The central idea behind probabilistic theories of causation is that "causes change the probability of their effects" [29]. A wall's shape partly determines the probability of hitting it, or hitting it in certain places, or in a certain manner. This is especially clear if we consider what might happen if, counterfactually, we were in a group of twenty. As group size increases, the chance that one of us will touch the wall rises significantly. See Figure 17. Since social interaction in groups in a single room is transitive, it is enough that anyone in the group touch the wall to warrant saying that everyone in the group interacts with (though not touches) the wall. When an outer person touches the wall, they will move away, thereby sending a shudder

of adaptive interaction through everyone as they jointly accommodate that person's positional change. This physical adaptation may be unconscious for some. But still present. If the group size shrinks, the probability of physical contact also shrinks. But from a probabilistic view, everyone interacts with walls in this physical way, even when they are alone; it is just that their chance of collision is rather low. This is the "probabilistically interact" at every moment view. It assumes that to interact with something we do not need to always overtly cause changes in it in response to its actions on us as long as periodically we do cause changes in it. The room partly determines the probability distribution of behavior.

If either of these arguments is compelling, it offers support to those architects who argue that we interact with walls when we navigate through a building, or when we collect in corners or alcoves, even when, as individuals, we never touch their surface.

Related cases are easy to invent. Suppose in our building the lights come on across the whole floor (as opposed to the local area where we are active) only when there are enough people over the course of an hour, regardless of where they are. Two of us arrive first. Only the nearby lights come on. Two more arrive. The threshold is reached and lighting for the whole floor comes on. We are all partial causes, so presumably we all have interacted with the building. Suppose now some of us leave before the threshold is reached. As others arrive within the hour, the lights come on throughout. Did the early leavers interact? Any one of us might have been the last straw, but only some of us were present to experience the effect. Must we all experience the output every occasion to be said to interact? Not if probabilistic interaction is acceptable. It is enough that causes change the probability of their effects. By coming into the space, we changed the probability that the lights would come on throughout. We don't have to receive feedback each occasion to be a cause and to be a probabilistic recipient – a sometimes experiencer – of an effect. If causal invariance were required then in artistic installations, where there is only a probability that a system will respond to any of our actions, we could not be said to interact with the system, except in those moments when it actively responds. Returning to the lighting case, if we never know our effect, we cannot explicitly intend to help cause the lights to go on. On the other hand, as long as we implicitly know there is a probabilistic connection then we are able to implicitly interact with the lights.

The network model of interaction deserves much more complete development. It can provide commentary on many of the difficult intuitions architects and others have about interacting with collectives, holistic structures like buildings, walls when we are in groups, and so on. It still falls short of the ecological model, though, in that it does not support reflexive interaction. We turn now to the last model of interaction, full blooded architectural interaction.

4.4 Interaction in Architectural Space

I have argued that only in systems containing social interfaces, or things just like them, do we find support for reflexive interaction. That is one of the special features that architects need bear in mind when thinking about interaction in and with buildings. Every sort of interaction can be found in architecture: direct manipulation, network interaction, ecological interaction, but it is co-present social interaction – a type of ecological interaction – that is what differentiates the architectural design problem from network and direct manipulation design problems.

Why is interaction sometimes reflexive in architectural space? Because the very act of moving oneself can create a social interface where none existed. When two people are close enough in physical space, especially physical space with walls, furniture and "props," their biochemistry changes – literally [27]. We don't have to speak or generate behavior that resembles device input to change our social interface; the change happens because we create and share bubbles of commonality: epistemic, value-based, experiential, and practical. By moving into range of another human, we become a small social group, a social bubble, something bigger than each of us, but composed of

us and our context. It follows that any change we make to ourselves is a change to the group bubble. We can act on our bubble from the inside. Changing our stance, physical distance, gaze, or fidgeting nervously, are all reflexive actions and they carry social meaning that can change the bubble.

To my knowledge, reflexive interaction is not yet a topic of study in HCI. It is a marker for a type of interactivity not subsumable in current HCI models. Overall, though, despite it being a well-defined attribute that can be used to distinguish interfaces, I suspect it will not be as important or interesting to designers as the idea of embodied interaction and niche-style ecological interfaces, which refer to an entire style of interaction. If there is gold in this inquiry here is where it lies. Why? Because architectural intuitions about interface and interaction raise hard questions about how people share physical space. There is so much we don't know about how people share their social, physical, and epistemic bubbles. How they co-create a space, or a place. How they see space as a resource to share, to coordinate activity in, to extend their mind into and to use in cognitive ways [42]. Indeed, what is the space of shared activity [66]? Yet, despite our conceptual naiveté, this embodied sharing is the very thing driving the new model of interaction; it constitutes the situated aspects of *being in a place* that context-aware systems cannot recover or recreate.

Here is an example of a bubble that has all the aspects we need to think about. To get the possibilities for interaction that being in this bubble offers you, you have to be an embodied human in this very space.

When a mother clothes her child the two participate in an intensely personal joint activity. The mother without intending to, is teaching her child about coordination, cooperation, and shared attention. Cartesian space is far from the mind and brain of each as spatial points of interest are jointly defined by the two as they touch and move, together coping with the awkwardness of dressing. There is an odd sort of joint first-person perspective [83, 50] – a shared peripersonal space [10]. To clothe oneself or another, it is necessary to comply with shapes and forms. Mother and child choreograph their movements to overcome the hassles of putting limbs in sleeves, and socks on feet. For the participants, space is grasped in an embodied manner where both parties understand what is happening in a highly indexical and situational manner, often using non-conceptual skills [20]. The two construct an interdependent social reality. Those outside the duet can appreciate but not grasp that constructed social reality in the same way owing to the neural consequences of sharing peripersonal space [75] and the feeling of personal involvement [75]. This social reality also includes a profound empathy for the experience of the other [65, 76]. People share experience almost as much as they share activity. To participate in these special activities, you have to be in close quarters with another, grasp shared peripersonal space and develop feelings of involvement. These are hard to digitally recreate.

The embodied and situated nature of dressing highlights the complex way space is understood by humans. Our knowledge of "where" is often grasped through relations defined by our bodies. Someone snaps their fingers. Our knowledge of where the sound comes from is not initially encoded in objective cartesian space. We encode it relationally; our knowledge may be better understood by our capacities to orient, to look in the right place, to be primed for additional sound from "there" [56]. This same type of indexical, embodied knowledge holds for my knowing where I am in relation to you, or my knowing where you are looking [15], knowing how you are moving your limbs and how I am moving mine. When two people are together, they assume that the other roughly hears what they hear, and each knows the other assumes that too. It is recursive. This is the sense in which participants jointly create an epistemic bubble that goes beyond what can be explicitly articulated in common ground[11] because it includes non-conceptual elements. Simon Penny has a thoughtful account of how non-conceptual, sensori-motor interaction lies at the heart

---

[11]See Clarke's original work on common ground (1995) and extensions by Klein et al. (2005).

of the best new media interactive installations [70]. The interactive possibilities that come from sharing a bubble without another person are not readily recreated.

A second bubble has to do with experience and empathy [6]. When we stand beside a friend, we likely share much of the same experience. We feel the same ambient temperature, we smell the same smells, see the same colors and things around us. The list is longer than you might think. That people differ at times and discuss their individual reactions is more an indication that we share so much that our differences are noteworthy. Sharing experience is possible in part because we engage the world through active and interactive perception [24]. We are not passive, our grasp of things around us is through sensori-motor interaction [18]. So not only do we encounter a similar world because we are a few feet from each other; we encounter it because in all likelihood we are participating in some form of joint activity that shapes a salience structure of the world around us [63, 67]. Once in this bubble unique interactive possibilities open up.

A third bubble, also related to experience, is about valuation [38]. People have different tastes and preferences, but our differences are drawn up against a background of similarity. Wittgenstein spoke of the necessity for intersubjective sense-making bottoming out in our common "form of life" [85]. Without knowing what it's like to be human, wanting and liking what humans like, feeling what humans feel, seeing the world through senses like our own, how could we establish the sort of deep embodied experiential common ground, we need to understand each other? Again, being inside this bubble opens up interactive possibilities that differ from those outside the bubble.

To date, our digital creations do not participate in these social, co-agency bubbles where interaction is far more complex. The simplest reason is embodiment. Motor understanding of space – shared space – and the haptic feel of materials and structure in local space are necessary for full participation in bubbles. Joint activity often happens in joint space and sometimes this requires discovering that we share peripersonal space – it's easy to collide. We need to fine tune our coordination with others, and this requires acquired sensitivities. Imagine two people juggling or carrying a couch up the stairs. The timing of force variation depends on our bodies and also on the spatial properties of the material mediators we need for embodied interaction. Right now, there are limitations on how successfully digital substitutes can recreate the sense of embodied understanding that spatial co-presence gives rise to.

This body-based understanding is not reducible to intellectual understanding of problem space, task structure, and sub-goal structure. There are other things going on that have to do with touch, timing, force, and constructing real-time understanding of others. Architects have a deep responsibility to understand how humans jointly interact. In large halls, they can subtly change how involved people feel by changing how visible others are and how easy it is to see the effect of one's single voice or body on others. There is a reason fascist and brutalist architecture is so cavernous: it overwhelms individuals, reinforcing the idea that the state is more important than the single citizens who jointly constitute it. Does HCI face similar possibilities if it operates with a shallow understanding of the psychology of human co-presence?

5 CONCLUSION

I have argued that HCI has pushed our notions of interface and interactivity far beyond direct manipulation. With context-aware systems, we need to think about network interaction where interaction can be implicit – invisible to the agent – it can be probabilistic, it can hold at the group level, and it may be bewilderingly complex. Despite this rise in complexity, interaction in network models still resembles direct manipulation in being symmetric and transitive. But the variety of ways interaction can be transitive and how transitivity works in networks is far more complicated than anything we find in object-based interaction, the direct manipulation model of the classical account. It is also significant that in network interaction, agents oftentimes are not only unaware

they are interacting, they may not have been the initiators of interaction. The role of agency, intention, and control may have to be rethought.

I argued, further, that network models and direct manipulation models of interaction do not support interaction that is reflexive: cases analogous to a person acting on themselves, such as laughing at oneself, scratching oneself, talking to oneself. They are not reflexive because if there is a digital boundary between human and other, humans must interact with the digital other through input sensors and devices. By definition, humans are separate from those digital things. There is an interface at the digital divide. This separation between us and digital systems means that when we close our eyes, or when we talk to ourselves – both being actions that are reflexive – we do not engage the digital system in a reflexive manner. The system, if it is context aware, may record us, it may react to us, but our interaction with it is mediated by its sensors.

This implies that we can interact with ourselves reflexively but with the system only mediately. The one way we can reflexively interact with something other than ourselves is if we are an integral part of the very system we interact with. In such cases, a simple reflexive action such as closing one's eyes, would be a reflexive interaction with something beyond us just if we are part of that system. Only then would self-directed actions be actions on the whole system. In duets, a move by one person has an effect on the other, and simultaneously on the duet as a whole. No tools, no mediation, no action through an input device. In co-present social interaction, we are an intrinsic part of the whole, this often happens in an interface provided by the built environment. I suggested that architects have a line on social interaction and its role in design that is not yet standard in HCI. This makes it interesting. The metaphor of this interaction is of organisms in a niche: when we interact together in a building, we change the properties of our jointly created niche in the process of interacting with each other.

Metaphors aside, what makes face-to-face social interaction interesting is that we have a special relation to those we socially interact with. Specifically, we are sensitized to what others are feeling, registering, valuing, and doing. This is a consequence of being embodied humans evolved for social interaction. When we act jointly, we coordinate in a mutually created space conceptualized with situationally specific concepts. These concepts may be jointly created and ad hoc [5], a type of shared epistemic bubble. Further, our grasp of many of the things in this bubble are sensori-motor based and non-conceptual. This makes it hard for AI and even robots to socially interact with us in a full-blooded manner. We can interact with these inventions in a direct manipulation way, or a network way. We can even interact in weakly social ways, exchanging words, working on projects together. But there is something missing: empathy, shared understanding of momentary experience, shared valuation, a sense of place. It will be interesting to see how architects change their designs when they expect humans to be interacting with robots much more. My own guess is that there will be less spatial intrigue in our built spaces and fewer nooks for social interaction.

Owing to the more embodied/embedded view of humans we find in architectural thinking, it would be surprising if there were not significant differences in the way architects and HCI designers approach their design problems. I mentioned six at the outset and see these as reflecting different requirements each field has faced historically. As the fields interpenetrate, we may predict new metaphors of interaction and interface to emerge, to the benefit of both.

REFERENCES

[1] R. Adolphs. 2003. Cognitive neuroscience: Cognitive neuroscience of human social behaviour. *Nature Reviews Neuroscience* 4, 3 (2003), 165.

[2] H. S. Alavi, E. Churchill, and D. Lalanne. 2017. The Evolution of Human-Building Interaction: An HCI Perspective PREFACE (No. ARTICLE, pp. 3–6). Interaction Design & Architectures.

[3] C. Alexander. 1977. *A Pattern Language: Towns, Buildings, Construction.* Oxford University Press.

[4] M. A. Arbib, J. B. Bonaiuto, S. Jacobs, and S. H. Frey. 2009. Tool use and the distalization of the end-effector. *Psychological Research PRPF* 73, 4 (2009), 441–462.

[5] L. Barsalou. 1983. Ad hoc categories. *Memory and Cognition* 11, 3 (1983), 211–227.

[6] B. C. Bernhardt and T. Singer. 2012. The neural basis of empathy. *Annual Review of Neuroscience* 35 (2012), 1–23.

[7] M. J. Bitner. 1992. Servicescapes: The impact of physical surroundings on customers and employees. *Journal of marketing*, 56, 2 (1992), 57–71.

[8] J. Blom. 2000. Personalization: A taxonomy. In *Proceedings of the Extended Abstracts on Human Factors in Computing Systems.* ACM, 313–314.

[9] P. Bourdieu. 2017. Habitus. In *Habitus: A Sense of Place.* Routledge, 59–66.

[10] C. Brozzoli, G. Gentile, L. Bergouignan, and H. H. Ehrsson. 2013. A shared representation of the space near oneself and others in the human premotor cortex. *Current Biology* 23, 18 (2013), 1764–1768.

[11] R. Buchanan. 1992. Wicked problems in design thinking. *Design Issues* 8, 2 (1992), 5–21.

[12] V. Caggiano, L. Fogassi, G. Rizzolatti, P. Thier, and A. Casile. 2009. Mirror neurons differentially encode the peripersonal and extrapersonal space of monkeys. *Science* 324, 5925 (2009), 403–406.

[13] S. K. Card, J. D. Mackinlay, and G. G. Robertson. 1991. A morphological analysis of the design space of input devices. *ACM Transactions on Information Systems* 9, 2 (1991), 99–122.

[14] L. A. Carlson-Radvansky and D. E. Irwin. 1993. Frames of reference in vision and language: Where is above? *Cognition* 46, 3 (1993), 223–244.

[15] M. Chita-Tegmark. 2016. Social attention in ASD: A review and meta-analysis of eye-tracking studies. *Research in Developmental Disabilities* 48 (2016), 79–93.

[16] A. Cussins. 2002. Experience, thought and activity. In *Essays on Nonconceptual Content*, Y. H. Gunther (Ed.). MIT Press, 133–163.

[17] N. S. Dalton, H. Schnädelbach, M. Wiberg, and T. Varoudis. 2016. *Architecture and Interaction.* Springer, Cham, 978–3.

[18] F. P. De Lange, M. Spronk, R. M. Willems, I. Toni, and H. Bekkering. 2008. Complementary systems for understanding action intentions. *Current Biology* 18, 6 (2008), 454–457.

[19] P. Dourish. 2004. *Where the Action is: The Foundations of Embodied Interaction.* MIT press.

[20] G. Evans. 1982. *The Varieties of Reference.* Oxford University Press.

[21] J. Feldman. 2003. The simplicity principle in human concept learning. *Current Directions in Psychological Science* 12, 6 (2003), 227–232.

[22] A. T. Friedman. 2006. *Women and the Making of the Modern House: A Social and Architectural History.* Yale University Press.

[23] C. D. Frith and U. Frith. 2008. Implicit and explicit processes in social cognition. *Neuron* 60, 3 (2008), 503–510.

[24] V. Gallese. 2003. The roots of empathy: The shared manifold hypothesis and the neural basis of intersubjectivity. *Psychopathology* 36, 4 (2003), 171–180

[25] T. Gu, H. K. Pung, and D. Q. Zhang. 2005. A service-oriented middleware for building context-aware services. *Journal of Network and Computer Applications* 28, 1 (2005), 1–18.

[26] B. Hardian, J. Indulska, and K. Henricksen. 2008. Exposing contextual information for balancing software autonomy and user control in context-aware systems. In *Proceedings of the Workshop on Context-Aware Pervasive Communities: Infrastructures, Services and Applications.*

[27] R. Hari, L. Henriksson, S. Malinen, and L. Parkkonen. 2015. Centrality of social interaction in human brain function. *Neuron* 88, 1 (2015), 181–193.

[28] M. Heidegger. 1962. *Being and Time. 1927.* (Trans. John Macquarrie and Edward Robinson). Harper, New York.

[29] Hitchcock Christopher. Probabilistic Causation. In *The Stanford Encyclopedia of Philosophy (Fall 2018 Edition)*, E. N. Zalta (Ed.). Retrieved from <https://plato.stanford.edu/archives/fall2018/entries/causation-probabilistic/>.

[30] N. P. Holmes and C. Spence. 2004. The body schema and multisensory representation (s) of peripersonal space. *Cognitive Processing* 5, 2 (2004), 94–105.

[31] K. Hornbæk and A. Oulasvirta. 2017. What is interaction? In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 5040–5052.

[32] E. Hornecker. 2011. The role of physicality in tangible and embodied interactions. *Interactions* 18, 2 (2011) 19–23.

[33] E. Hornecker and J. Buur. 2006. Getting a grip on tangible interaction: A framework on physical space and social interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 437–446.

[34] R. A. Howard. 2012. *Dynamic Probabilistic Systems: Markov Models*, Vol. 1. Courier Corporation.

[35] L. E. Janlert and E. Stolterman. 2017a. *Things That Keep Us Busy: The Elements of Interaction*. MIT Press.

[36] L. E. Janlert and E. Stolterman. 2017b. The meaning of interactivity—Some proposals for definitions and measures. *Human–Computer Interaction* 32, 3 (2017b), 103–138.

[37] W. Ju and L. Leifer. 2008. The design of implicit interactions: Making interactive systems less obnoxious. *Design Issues* 24, 3 (2008) 72–84.

[38] J. W. Kable and P. W. Glimcher. 2007. The neural correlates of subjective value during intertemporal choice. *Nature Neuroscience* 10, 12 (2007), 1625.

[39] I. Kant. 1993. *Grounding for the Metaphysics of Morals: With on a Supposed Right to Lie Because of Philanthropic Concerns*. Hackett Publishing.

[40] O. Khatib. 1987. A unified approach for motion and force control of robot manipulators: The operational space formulation. *IEEE Journal on Robotics and Automation* 3, 1 (1987), 43–53.

[41] D. Kirsh. 2013. Embodied cognition and the magical future of interaction design. *ACM Transactions on Computer-Human Interaction* 20, 1 (2013), 3.

[42] D. Kirsh. 1995. The intelligent use of space. *Artificial Intelligence* 73, 1–2 (1995), 31–68.

[43] G. Knoblich, S. Butterfill, and N. Sebanz. 2011. Psychological research on joint action: Theory and data. In *Psychology of Learning and Motivation*, Vol. 54. Academic Press, 59–101.

[44] J. Krueger. 2014. Affordances and the musically extended mind. *Frontiers in Psychology*, 4, 1003.

[45] S. Kühn, M. Brass, and P. Haggard. 2013. Feeling in control: Neural correlates of experience of agency. *Cortex* 49, 7 (2013), 1935–1942.

[46] B. Latour. 2005. *Reassembling the Social: An Introduction to Actor-Network-Theory*. Oxford University Press.

[47] T. Lozano-Perez. 1990. Spatial Planning: A Configuration Space Approach. In *Autonomous Robot Vehicles*, Ingemar J. Cox and Gordon T. Wilfong (Eds.). Springer, New York, NY, 259–271.

[48] G. F. Luger. 2005. *Artificial Intelligence: Structures and Strategies for Complex Problem Solving*. Pearson Education.

[49] W. L. MacDonald. 2002. *The Pantheon: Design, Meaning, and Progeny*. Harvard University Press.

[50] L. Maister, F. Cardini, G. Zamariola, A. Serino, and M. Tsakiris. 2015. Your place or mine: Shared sensory experiences elicit a remapping of peripersonal space. *Neuropsychologia* 70 (2015), 455–461.

[51] J. Malpas. 2018. *Place and Experience: A Philosophical Topography*. Routledge.

[52] A. Maravita and A. Iriki. 2004. Tools for the body (schema). *Trends in Cognitive Sciences* 8, 2 (2004), 79–86.

[53] D. Marr. 1976. Early processing of visual information. *Philosophical Transactions of the Royal Society London B* 275, 942 (1976), 483–519.

[54] D. Marr and E. Hildreth. 1980. Theory of edge detection. *Proceedings of the Royal Society B* 207, 1167 (1980), 187–217.

[55] M. Martel, L. Cardinali, A. C. Roy, and A. Farnè. 2016. Tool-use: An open window into body representation and its plasticity. *Cognitive Neuropsychology* 33, 1–2 (2016), 82–101.

[56] J. J. McDonald, W. A. Teder-Sälejärvi, and S. A. Hillyard. 2000. Involuntary orienting to sound improves visual perception. *Nature* 407, 6806 (2000), 906.

[57] M. Merleau-Ponty. 1945/1962. *Phenomenology of Perception*. C. Smith (trans.). Routledge, New York and London. Originally published in French as Phénoménologie de la Perception.

[58] J. W. Moore, D. Middleton, P. Haggard, and P. C. Fletcher. 2012. Exploring implicit and explicit aspects of sense of agency. *Consciousness and Cognition* 21, 4 (2012), 1748–1753.

[59] M. J. Muller, J. Freyne, C. Dugan, D. R. Millen, and J. Thom-Santelli. 2009. Return on contribution (ROC): A metric for enterprise social software. In *Proceedings of the European Conference on Computer-Supported Cooperative Work*. Springer, London, 143–150.

[60] P. Mundy and L. Newell. 2007. Attention, joint attention, and social cognition. *Current Directions in Psychological Science* 16, 5 (2007) 269–274.

[61] B. Myers, S. E. Hudson, and R. Pausch. 2000. Past, present, and future of user interface software tools. *ACM Transactions on Computer-Human Interaction* 7, 1 (2000), 3–28.

[62] H. A. Simon and A. Newell. 1972. *Human Problem Solving*, Vol. 104, No. 9. Prentice-Hall, Englewood Cliffs, NJ.

[63] A. Nowak, R. R. Vallacher, M. Zochowski, and A. Rychwalska. 2017. Functional synchronization: The emergence of coordinated activity in human systems. *Frontiers in Psychology* 8 (2017).

[64] I. Newton, A. Motte, and N. W. Chittenden. 1850. *Newton's Principia: The Mathematical Principles of Natural Philosophy*. George P. Putnam.

[65] K. N. Ochsner, K. Knierim, D. H. Ludlow, J. Hanelin, T. Ramachandran, G. Glover, and S. C. Mackey. 2004. Reflecting upon feelings: An fMRI study of neural systems supporting the attribution of emotion to self and other. *Journal of Cognitive Neuroscience* 16, 10 (2004), 1746–1772.

[66] E. Pacherie. 2011. The Phenomenology of Joint Action: Self-Agency versus Joint Agency. In *Joint attention: New Developments in Psychology, Philosophy of Mind, and Social Neuroscience*, Seemann Axel (Ed.). MIT Press, 2011.

[67] T. Parr and K. J. Friston. 2017. Working memory, attention, and salience in active inference. *Scientific Reports* 7, 1 (2017), 14678.

[68] J. Pearl and D. Mackenzie. 2018. *The Book of Why: The New Science of Cause and Effect*. Basic Books.

[69] P. B. Pearman, A. Guisan, O. Broennimann, and C. F. Randin. 2008. Niche dynamics in space and time. *Trends in Ecology & Evolution* 23, 3 (2008), 149–158.

[70] S. Penny. 2017. *Making Sense: Cognition, Computing, Art, and Embodiment*. MIT Press.

[71] U. J. Pfeiffer, K. Vogeley, and L. Schilbach. 2013. From gaze cueing to dual eye-tracking: Novel approaches to investigate the neural correlates of gaze in social interaction. *Neuroscience & Biobehavioral Reviews* 37, 10 (2013), 2516–2528.

[72] F. H. Previc. 1998. The neuropsychology of 3-D space. *Psychological Bulletin* 124, 2 (1998) 123.

[73] L. J. Raphael, G. J. Borden, and K. S. Harris. 2007. *Speech Science Primer: Physiology, Acoustics, and Perception of Speech*. Lippincott Williams & Wilkins.

[74] G. Ryle. 1949. *The Concept of Mind*. Hutchinson & Co. Ltd., London.

[75] L. Schilbach, A. M. Wohlschlaeger, N. C. Kraemer, A. Newen, N. J. Shah, G. R. Fink, and K. Vogeley. 2006. Being with virtual others: Neural correlates of social interaction. *Neuropsychologia* 44, 5 (2006) 718–730.

[76] M. Schulte-Rüther, H. J. Markowitsch, G. R. Fink, and M. Piefke. 2007. Mirror neuron and theory of mind mechanisms involved in face-to-face interactions: A functional magnetic resonance imaging approach to empathy. *Journal of Cognitive Neuroscience* 19, 8 (2007) 1354–1372.

[77] B. Shneiderman. 2010. *Designing the User Interface: Strategies for Effective Human-Computer Interaction*. Pearson Education, India.

[78] H. A. Simon. 1973. The structure of ill structured problems. *Artificial Intelligence* 4, 3–4 (1973), 181–201.

[79] M. Synofzik, G. Vosgerau, and A. Newen. 2008. Beyond the comparator model: A multifactorial two-step account of agency. *Consciousness and Cognition* 17 (2008), 219–239.

[80] M. Tomasello. 1995. Joint attention as social cognition. In *Joint Attention: Its Origins and Role in Development*, C. Moore and P. Dunham (Eds.), 103–130.

[81] G. Vallar and A. Maravita. 2009. Personal and extra-personal spatial perception. In *Handbook of Neuroscience for the Behavioral Sciences*, Gary Berntson and John Cacioppo (Eds.). John Wiley and Sons, Inc., 322–336.

[82] T. L. van Zuijen, V. L. Simoens, P. Paavilainen, R. Näätänen, and M. Tervaniemi. 2006. Implicit, intuitive, and explicit knowledge of abstract regularities in a sound sequence: an event-related brain potential study. *Journal of Cognitive Neuroscience* 18, 8 (2006), 1292–1303.

[83] K. Vogeley and G. R. Fink. 2003. Neural correlates of the first-person-perspective. *Trends in Cognitive Sciences* 7, 1 (2003), 38–42.

[84] B. Waber, J. Magnolfi, and G. Lindsay. 2014. *Workspaces That Move People*. Harvard Business Review.

[85] L. Wittgenstein. 1953. *Philosophical Investigations*. Basil Blackwell.

[86] K. Yun, K. Watanabe, and S. Shimojo. 2012. Interpersonal body and neural synchronization as a marker of implicit social interaction. *Scientific Reports* 2 (2012), 959.