

# POINTING, MUTUAL INTELLIGIBILITY, AND THE SEEING SUBJECT IN HCI

By Jonathan Zong

*“Users, who select data objects by looking and pointing, are simultaneously also the objects being seen, selected, and acted upon by computers.”*

Suggested citation:

Jonathan Zong, Pointing, Mutual Intelligibility, and the Seeing Subject in HCI. *Interface Critique* 4 (2022): 95–105.

DOI: <https://doi.org/10.11588/ic.2023.4.93413>

This article is released under a Creative Commons license (CC BY 4.0).

*This essay was originally contributed to the 2021 Computer Mouse Conference.*

A grainy video depicts a man sitting in an Eames lounge chair, facing a wall-sized projection screen. As he points at the screen, a cross-shaped cursor “tracks” where he points. The man issues a few voice commands, creating four symbols with distinct colors and shapes at various positions. When he points to a symbol and then to a new location, he says “put that ... there,” relocating the symbol to the new location. After the man executes a series of increasingly complex voice commands, the system malfunctions. “Ah, shit,” he says as the video ends.



Fig. 0: Chris Schmandt, “Put-That-There” (November 2, 1979), MIT Media Lab Speech Interface group video collection. Source: <https://www.youtube.com/watch?v=RyBEUyEtXQo>.

Created in 1979, the “Put-That-There” system was part of an MIT research project on ways to access and manipulate data spatially using pointing. “Put-That-There” exemplifies how the field of human-computer interaction (HCI) has constructed the human subject. The

researchers conceived of data – represented in the demo as circles, squares, and triangles – as “inhabiting a spatially definite ‘virtual’ world,”<sup>1</sup> which computer users could access through a multisensory technical apparatus. The researchers hoped to immerse users in an information environment where users could see and move data around. To make user interactions legible to the computer, researchers needed to grapple with questions about how to represent people in virtual space. The creators of “Put-That-There” aspired for users to think of data as objects they could sense, “bodied forth in vision, sound, and touch”<sup>2</sup>. But to do so, they also needed ways for computers to understand users computationally – as humans bodied forth in data.

This essay proceeds in two parts. In the first part, I situate HCI’s subject – the user – in conversation with prior theories about how visual media constructs seeing subjects. “Put-That-There” was designed according to theories in HCI about interaction as a feedback loop of perception and action between users and computers. Past theories in film and photography argued that the act of seeing establishes a strict spatial division between subject and object.<sup>3</sup> Being able to observe something in an image meant that the observer was not part of the image. I argue that interactivity com-

1 Richard Bolt, *Spatial Data-Management* (Cambridge, MA 1979), p. 12

2 Ibid.

3 Jonathan Crary, *Techniques of the Observer: On Vision and Modernity in the Nineteenth Century* (Cambridge, MA 1990).

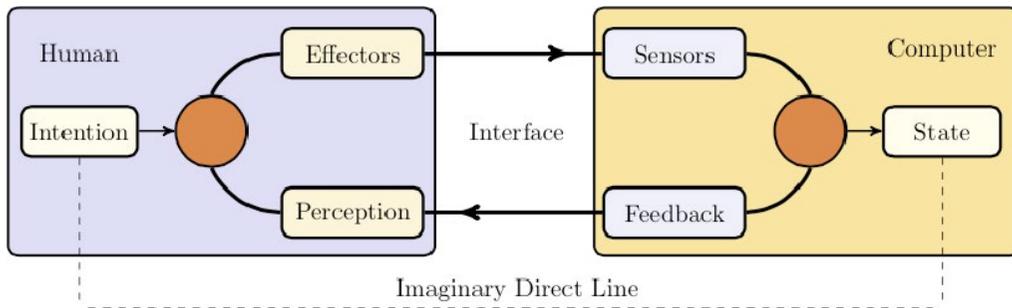


Fig. 1: HCI model of interaction. Source: <https://www.semanticscholar.org/paper/Empowerment-as-a-metric-for-Optimization-in-HCI-Trendafilov-Murray-Smith/3c37fcd2bb95f498b7c181c7e5180e9433d8ffcb/figure/0>.

plicates this strict division. In interactive systems, it is now possible for the user to act on visual representations of virtual objects.

In the second part, I dig into a specific way interactivity complicates this division. Interactivity reconfigures the relationship between subject and object: from a unidirectional relationship of observation to a bidirectional relationship of mutual intelligibility. By positioning the user within a feedback loop, HCI establishes symmetry between the user and the computer. Users who act on data are also *acted upon* by data. To make this argument, I give an account of some fundamental operations in interaction – including selection and identification – and suggest that they establish common perceptual ground between human and machine interlocutors. Pointing devices, such as the computer mouse, play an important role in enabling users to manipulate data. But because interaction is bidirectional, these same operations enable computers to manipulate people.

## Part I: Situating the User in the History of the Seeing Subject

### Modeling Users as Information Processors

Computers are containers of virtual worlds populated by data objects. As such, they can only perceive the external world through input devices such as computer mice, which translate physical actions into electronic signals. Similarly, they can only make virtual objects perceptible to human observers by creating sensory representations, using output devices like screens.

Computers sense the world through inputs and outputs, but HCI researchers have also conceptualized people as I/O machines. Influenced by cognitive science and cybernetics, the field theorizes interaction as a feedback loop between

a user and a system.<sup>4</sup> In this model, the user is essentially an information processing machine. The user has a sense input (e.g., eyes), a control (some cognitive map of their goals and intentions), and an articulatory output (e.g., the ability to move a computer mouse).

The term “user,” though seemingly referencing personhood, is best understood as the particular way HCI’s underlying theoretical framework constructs the subject. HCI researchers constructed this model in order to make the concept of a person operationalizable in computer systems. To be understood by machines, humans had to conform to a machine-like schema of input and output. As a result, Lasse Scherffig writes, “the human trained to perform in front of the computer became the model for the thinking human in general – a human acting as a computer”<sup>5</sup>. In order to perceive and act on data objects in the virtual world, people need to adopt the subject position of users – behaving in ways that allow them to become read as data themselves.

### How the Computer Sees Us

The idea that technologies rearrange how we think about human sense faculties is not new. Before scholars asked these questions about computers, there were similar, historically preceding debates about other forms of media such as

film. For instance, Jean-Louis Baudry argues that the technical systems and cultural practices that go into producing film (the cinematic apparatus) are not merely neutral, but have ideological effects that construct the spectator as a subject.<sup>6</sup> Because film viewers see through the perspective of a single monocular camera, and their body stays still while the camera seems to jump to different locations and times, theories of film have assumed a spectator that sees “with a single and immobile eye”<sup>7</sup>. Just as HCI theorists argue that users’ access to virtual worlds is limited by the technical sensory apparatus available to computers, film theorists recognize the particular way that the camera, editing, and projection afford a limited way of experiencing cinematic worlds.

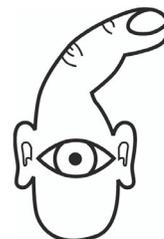


Fig. 2: “How the computer sees us.” Source: Dan O’Sullivan and Tom Igoe, *Physical Computing: Sensing and Controlling the Physical World with Computers* (Boston 2004).

HCI’s interaction model is continuous with these prior attempts to theorize how sociotechnical apparatuses

4 Lasse Scherffig, There Is No Interface (Without a User). A Cybernetic Perspective on Interaction. *Interface Critique Journal* 1 (2018): 58–80.

5 Ibid., p. 67.

6 Jean-Louis Baudry, Ideological Effects of the Basic Cinematographic Apparatus. *Film Quarterly* 28/2 (1974): 39–47.

7 Erwin Panofsky, *Perspective as Symbolic Form* (New York 1991), p. 29.

shape people's experiences. Baudry's "eye-subject" has been succeeded by, for instance, Dan O'Sullivan and Tom Igoe's illustration of "how the computer sees us" – as a single eye augmented with a single finger.<sup>8</sup> As bizarre as it looks, the eye-finger-subject is illustrative of the way the field of HCI thinks about the human sensorium in terms of interface modalities. The eye and ears represent the human perceptual capacities that computers often use to output data, by rendering it visible or audible. The single finger represents a primary way computers receive human input: through pointing, or through the mechanical actuation of mouse and keyboard buttons. The illustration lacks a mouth – perhaps the authors did not want to distinguish different mouth functions like speaking and tasting – but the fact that the illustration is somewhat contrived is also the point. The idea of conforming a person's body to an apparatus is necessarily contrived.

### Positioning the Body in Relation to Data

Like "Put-That-There," a *camera obscura* is a room with a person inside. Light from outside the room passes through a pinhole into the otherwise dark space and projects an inverted image opposite the pinhole. As a predecessor to contemporary photographic technologies, the *camera obscura* has been an important case for theorizing vision. In *Techniques*

*of the Observer*, Jonathan Crary explains how the camera enforces a spatial division between subject and object: "the camera obscura *a priori* prevents the observer from seeing his or her position as part of the representation"<sup>9</sup>. That is, if one is situated inside the camera apparatus and able to observe the visual image captured from outside, one cannot be an object represented in the image – and vice versa. Crary notes that "the body then is a problem the camera could never solve except by marginalizing it into a phantom in order to establish a space of reason"<sup>10</sup>.

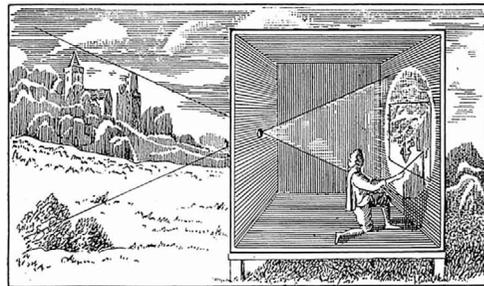


Fig. 3: Illustration of a camera obscura. Source: <https://magazine.artland.com/agents-of-change-camera-obscura/>.

Graphical user interfaces, and the broader project of interactivity in HCI, complicate this strict spatial division of subject and object. The "media room"<sup>11</sup>, as the MIT researchers called the setting of "Put-That-There," is formally similar to a *camera obscura* – an enclosed technical apparatus containing both an observer and

<sup>8</sup> Dan O'Sullivan and Tom Igoe, *Physical Computing: Sensing and Controlling the Physical World with Computers* (Boston 2004).

<sup>9</sup> Crary, *Techniques of the Observer*, p. 41.

<sup>10</sup> Ibid.

<sup>11</sup> Bolt, *Spatial Data-Management*, p. 12.

an image projected onto a wall. The chair at the center of the room might draw comparisons to the cinematic spectator's seat, immobilizing the user. But in the space of the graphical interface, the presence of the user is represented by a cursor. The cross-shaped cursor in "Put-That-There" tracks the intersection of the imaginary line extending out from the user's index finger with the image plane on the wall. Its jittery movement as the user's liveness keeps their hand continually in motion visualizes some element of what Crary calls a "spatial and temporal simultaneity of human subjectivity and objective apparatus"<sup>12</sup>. The cursor is a data object, and is positioned inside the virtual space of the screen just like other data objects; yet it represents and is controlled by the user. Unlike the *camera obscura*, the user sees themselves within the image despite occupying a separate space from the objects being represented. Reading "Put-That-There" in the historical lineage of photography and film helps us recognize the interactive cursor as a site where the computer user departs from prior constructions of the seeing subject.

## Part II: Establishing Mutual Intelligibility through Interaction

### Selection as a Building Block of Interaction

Cursors are fundamental to human-computer interaction because they allow the user to identify which data objects, out of all the objects in their field of perception, to act upon. In computer science, "selection" refers to an operation for querying a subset of data from a larger dataset. A selection is defined using a logical restriction on data attributes that evaluates to true or false. In the below example, the full "Person" dataset in the left column contains a list of 5 people. The right column contains a selection of people whose age is greater than or equal to 34. The "is greater than or equal to" logical restriction neatly cleaves the original dataset into two subsets: one which satisfies the restriction, and one which does not. Conventionally, we say that those 34 and older are *included* in the selection and the others are *excluded*.

When a user of "Put-That-There" points at a shape and says the word "that," they are specifying a selection that includes the indicated data object. The selection is defined using an implicit logical restriction: data points with a spatial position equal to that of the cursor. Interface designers leverage pointing as a way to

---

<sup>12</sup> Crary, *Techniques of the Observer*, p. 41.

select data objects by their position in space. To enable pointing-based selection, interfaces often spatialize data that is not necessarily inherently spatial. In the physical world, no two objects can occupy the same space at the same time. By designing interfaces such that this property also holds, spatial position can be made to serve as an identity.

Person		
Name	Age	Weight
Harry	34	80
Sally	28	64
George	29	70
Helena	54	54
Peter	34	80

$\sigma_{\text{Age} \geq 34}(\text{Person})$		
Name	Age	Weight
Harry	34	80
Helena	54	54
Peter	34	80

Fig. 4: An example of selection. Source: [https://en.wikipedia.org/wiki/Selection\\_\(relational\\_algebra\)](https://en.wikipedia.org/wiki/Selection_(relational_algebra)).

### Human-Computer Interaction as Joint Attention

Because selection allows users and computers to refer to objects in the same environment, it creates the common context that makes interaction possible. In *Plans and Situated Actions*, Lucy Suchman writes that “interaction, or communication – I’ll use the two interchangeably – turns on the extent to which my words and actions and yours are mutually intelligible”<sup>13</sup>. For Suchman, human-computer interaction is only made possible

by establishing a common ground for perception and action. When a user of “Put-That-There” points at a data object using the cursor, the computer is able to use the resulting selection as a proxy for understanding the user’s intent to apply subsequent voice commands to the selected object.

Some scholars have theorized attention as a selection of features out of a perceptual environment for the purpose of informing action.<sup>14</sup> In the interaction loop, because the user and a computer reference the same selection, they can be understood as attending to the same features of the virtual environment. When people communicate in physical space, pointing often expresses an invitation to joint attention – inviting others to redirect their attention to an indicated location. It might be a foundational way of expressing such an invitation – for instance, babies learn to point before they can speak.<sup>15</sup> Pointing at objects using cursors similarly facilitates joint attention between the user and the computer.

Where previously the seeing subject was often conceived of as a passive observer of the world, the user and the computer are constructed as equal, active participants within a feedback loop. Philosophers have theorized joint attention as a form of collective intentionality, which figures the world as “perceptu-

14 Wayne Wu, Attention as Selection for Action, in: *Attention: Philosophical and Psychological Essays*, eds. Christopher Mole, Declan Smithies and Wayne Wu (Oxford 2011), pp. 97–116.

15 Sotaro Kita, Pointing: A Foundational Building Block of Human Communication, in: *Pointing: Where Language, Culture, and Cognition Meet*, ed. Sotaro Kita (Mahwah, NJ 2003), pp. 1–8.

13 Lucy Suchman, *Plans and Situated Actions: The Problem of Human-Machine Communication* (Cambridge 1987), p. 3.

ally available for a plurality of agents ... [establishing] a basic sense of common ground on which other agents may be encountered as potential cooperators”<sup>16</sup>. Because interaction is a feedback loop, human attention and action is necessarily followed by machine attention and action.



Fig. 5: “Pointer Pointer” (2012) by Studio Moniker, an interactive website that surfaces an image of a person pointing to the location of your cursor. Source: <https://pointerpointer.com/>.

### Biometrics as Selection over Users

Where selections initiated by users allow humans to focus computer attention for the purpose of interaction, selections initiated by computers are increasingly used as a way to focus computers’ gazes upon people – for computers to determine who is human. Users perform selection through pointing, typing, and other forms of motion. But in addition to specifying selection, these movements often generate additional data as software records measurements of activity during everyday use – often without users’ knowledge. Logs of mouse move-

ments, records of keystrokes, amount of time spent on a webpage; Melissa Gregg compares this excess data to sweat, which “literalizes porosity” and is a “means by which the body signals its capacity to ‘affect and be affected’”<sup>17</sup>. Biometric data collected in the background of computer use is then used to select, differentiate, identify, and classify people – and to eventually target them more effectively.

Biometric profiles exemplify the process through which computers model and process humans as data objects – more precisely, objects assembled from the accumulation of data. For instance, proponents of digital psychiatry claim to be able to use biometric signals to diagnose and pathologize.<sup>18</sup> As a result, a market for biometric software that collects large amounts of data on key press timing has emerged in digital health-care. This software models the user as a collection of behavioral facts. It defines logical criteria through which computers can define selections of users on the basis of these facts. As anthropologist Beth Semel notes, “diagnoses also operate as vectors of social control” as people are partitioned into categories of well and unwell, deserving and undeserving of

16 David P. Schweikard and Hans Bernhard Schmid, Collective Intentionality, in: *The Stanford Encyclopedia of Philosophy* (2020).

17 Melissa Gregg, Inside the Data Spectacle. *Television & New Media* 16 (2014), pp. 37–51; <https://journals.sagepub.com/doi/full/10.1177/1527476414547774>, access: September 15, 2022.

18 Jonathan Zong and Beth Semel, Form, Content, Data, Bodies: Jonathan Zong and Beth Semel on Biometric Sans. *Somatosphere* (February 9, 2021); <http://somatosphere.net/2021/form-content-data-bodies.html/>, access: September 15, 2022.

clinical attention.<sup>19</sup> Inclusion and exclusion in these selection criteria consequently affect people's ability to navigate digitally-managed healthcare systems. Users, who select data objects by looking and pointing, are simultaneously also the objects being seen, selected, and acted upon by computers.



Fig. 6: "Biometric Sans" (2018) by Jonathan Zong, an experimental typography system which elongates letterforms in response to the typing speed of the individual. Source: <https://jonathanzong.com/blog/2020/05/31/biometric-sans-and-public-display-embodied-writing-in-the-age-of-data>.

## Conclusion: The One Divides into Two

Selection and identification – in other words, pointing things out – form the basis of human-computer interaction. These operations facilitate the feedback loop that is central to the field's understanding of the user as a subject. These

19 Beth Semel, *The Body Audible: From Vocal Biomarkers to a Pnenology of the Throat*. *Somatosphere* (September 21, 2020); <http://somatosphere.net/2020/the-body-audible.html/>, access: September 15, 2022.

operations are really the same operation of differentiation: to identify or select an object, one must articulate criteria that differentiate that object from others. Identifying a single object out of many requires criteria of inclusion and exclusion that cleave the space of possible referents into a binary partition – "this" and "not that".

This act of setting boundaries and creating binaries is fundamentally digital. Anthropologist Gregory Bateson defines the elementary unit of information as "a difference which makes a difference"<sup>20</sup>. Digital computers encode information in bits, which are basic units of differentiation. Alexander Galloway defines the digital as "the one divides into two," or more precisely, "any mode of representation rooted in individually separate and distinct units"<sup>21</sup>. Galloway's definition helps us see photography and film as predecessors to digital computers, because those media established subject and object as distinct binary units. Just as 0 can never be 1, the seeing subject could never be an object of representation. Drawing binaristic distinctions of inclusion and exclusion, interior and exterior, virtual and actual – these form the basis of working with computational media.

Yet, in conceiving of interaction as a feedback loop, HCI has constructed the user at various times as both subject and object of interaction. Where the re-

20 Gregory Bateson, *Form, Substance, and Difference*, in: *Steps to an Ecology of Mind* (San Francisco 1972), pp. 448–466.

21 Alexander Galloway, *Something About the Digital* (2015); <http://cultureandcommunication.org/galloway/something-about-the-digital>, access: August 2, 2022.

lationship between the subject-object binary was once a strict division, the two are cast by interaction as a set of roles that are adopted in turn. A user might select data objects, then be selected as a data object in turn. The user points, and the computer points back. Pointing is possible because difference exists, because there is something *else* to point at. Pointing is digital in this sense, and in the more literal sense that it happens using “the hand and its digits”<sup>22</sup>. However, Scherffig observes that “interaction fuses bodily activity and perception into one experience”<sup>23</sup>. The pointing finger is inextricable from the seeing eye. In this fusion, I see an attempt by human-computer interaction to work against the dominant tendency of digitality – to reconstitute the one from the two.

*Thank you to Arvind Satyanarayan, Haley Schilling, Kathleen Ma, Alan Lundgard, Crystal Lee, Drew Wallace, Geoffrey Litt, and members of the MIT Visualization Group for feedback on drafts of this piece! Thank you to Emma Rae Bruml for the invitation to contribute to the Computer Mouse Conference!*

---

22 Lisa Nakamura, Indigenous Circuits: Navajo Women and the Racialization of Early Electronic Manufacture. *American Quarterly* 66/4 (2014), pp. 919-941.

23 Lasse Scherffig, *Feedbackmaschinen. Kybernetik und Interaktion*. Dissertation, Academy of Media Arts Cologne (2017); <http://lassescherffig.de/publications/books/feedbackmaschinen-kybernetik-und-interaktion/>, access: August 2, 2022.

## References

- Bateson, Gregory**, Form, Substance, and Difference, in: *Steps to an Ecology of Mind* (San Francisco: Chandler Publishing Company, 1972), pp. 448–466.
- Baudry, Jean-Louis**, Ideological Effects of the Basic Cinematographic Apparatus. *Film Quarterly* 28 (1974): 39–47.
- Bolt, Richard**, *Spatial Data-Management* (1979).
- Crary, Jonathan**, *Techniques of the Observer: On Vision and Modernity in the Nineteenth Century* (Cambridge, MA: MIT Press, 1990).
- Galloway, Alexander**, Something About the Digital (2015); <http://cultureandcommunication.org/galloway/something-about-the-digital>, access: August 2, 2022.
- Gregg, Melissa, Inside the Data Spectacle. *Television & New Media* 16 (2014): 37–51.
- Kita, Sotaro**, Pointing: A Foundational Building Block of Human Communication, in: *Pointing: Where Language, Culture, and Cognition Meet*, ed. Sotaro Kita (Mahwah, NJ: Erlbaum, 2003), pp. 1–8.
- Nakamura, Lisa**, Indigenous Circuits: Navajo Women and the Racialization of Early Electronic Manufacture. *American Quarterly* 66/4 (2014): 919–941.
- O’Sullivan, Dan and Tom Igoe**, *Physical Computing: Sensing and Controlling the Physical World with Computers* (Boston: Thomson, 2004).
- Panofsky, Erwin**, *Perspective as Symbolic Form* (New York: Zone Books, 1991).
- Scherffig, Lasse**, *Feedbackmaschinen. Kybernetik und Interaktion*. Dissertation,

Academy of Media Arts Cologne (2017); <http://lassescherffig.de/publications/books/feedbackmaschinen-kybernetik-und-interaktion/>, access: August 2, 2022.

**Scherffig, L.** There Is No Interface (Without a User). A Cybernetic Perspective on Interaction. *Interface Critique Journal* 1 (2018): 58–80.

**Schweikard, David P. and Hans Bernhard Schmid.** Collective Intentionality. *The Stanford Encyclopedia of Philosophy* (2020).

**Semel, Beth.** The Body Audible: From Vocal Biomarkers to a Phrenology of the Throat, *Somatosphere* (September 1, 2020); <http://somatosphere.net/2020/the-body-audible.html/>, access: September 15, 2022.

**Suchman, Lucy.** *Plans and Situated Actions: The Problem of Human-Machine Communication* (Cambridge: Cambridge University Press, 1987).

**Wu, Wayne.** Attention as Selection for Action, in: *Attention: Philosophical and Psychological Essays*, eds. Christopher Mole, Declan Smithies, Wayne Wu (Oxford: Oxford University Press, 2011), pp. 97–116.

**Zong, Jonathan and Beth Semel.** Form, Content, Data, Bodies: Jonathan Zong and Beth Semel on Biometric Sans. *Somatosphere* (February 9, 2021); <http://somatosphere.net/2021/form-content-data-bodies.html/>, access: September 15, 2022.