THICK & THIN: “DIRECT MANIPULATION” & THE SPATIAL REGIMES OF HUMAN-COMPUTER INTERACTION

ABSTRACT
Consider a design trajectory, figured on one end by the screens of early command-line computer interfaces and ColecoVision’s “Donkey Kong” (1981), and on the other, by the more complex and finely rendered spaces depicted in “Tomb Raider III: The Adventures of Lara Croft” (1999) and Apple’s newly-released Mac OS X.

Lara Croft runs, jumps, tumbles, and blasts away at her opponents in visual fields that are more subtle and perspectivally sophisticated than those inhabited by Mario and “Donkey Kong.” The responses of the screen images to the user’s keyboard, gamepad, or joystick have been enormously enhanced, in both quickness and variety. But the fundamental spatial tropology—the tropology of space: abstract space, empty space, space that doesn’t get in the way of players or their agents on the other side of the glass—remains consistent, from the earliest to the most recent examples of both desktop computing interfaces and computer gaming. The conceptual and psychological commonplace that grounds play in the domains inhabited by Mario and Lara, and the principles of “direct manipulation” in the graphical user interface, is the assumption of a permeable field of agency, essentially free of substance or resistance, or marked only by the sorts of resistance that a more efficient game pad, a faster processor, or a more “intuitive” visual metaphor, may eliminate.

In this paper, I propouse that the “thin” spaces typical of the modern GUI and videogaming appear self-evident or “intuitive” to users and designers because they draw upon conventions of spatial thought that strategically foreclose traits of actual embodied encounters of human-computer interaction. It is desirable, I argue, to reconceive the forms of space commonly presupposed by the contemporary discourses of the GUI—to grasp these spaces materially, not as empty domains, open to the user’s purposive manipulations of objects sited within them, but rather as persistently impermeable, resistant—“thick”—spaces, in which objects are only imperfectly manipulated and incompletely detachable from the lived moment of the interaction.

KNOWLEDGE ON THE SCREEN
This is Donald Norman, from his 1994 CD-ROM for Voyager, Defending Human Attributes in the Age of the Machine:

Donald Norman describes the advantages of the GUI in his Defending Human Attributes in the Age of the Machine (1994), the Voyager Company. Used by permission.
Critics and designers of new media should be wary of the epistemological sleight of hand that makes this seem easy or obvious. To say that the GUI puts knowledge “on the screen” – a version of Norman’s signature distinction between “knowledge in the head” and “knowledge in the world” – may be appropriate for pragmatic analysis of the GUI’s dominance of desktop computing. Nonetheless, it leaves little room for critical thinking about the spatiality of the digital field or the conditions of knowledge it presumes, because it too narrowly circumscribes the terms of investigation.

“You can just look,” Norman promises of the GUI. But just looking is, strictly speaking, impossible for the intractably inconsistent consciousness we summarily describe as the “user.” Looking will always be caught between moments of seeing and not seeing, bracketed and deformed by historical, cultural, and technical practices that determine the viewer’s grasp of what it means to see anything at all. Before we can carefully discuss the string of something called knowledge “on the screen,” we need to investigate the assumption, widely held by designers of human-computer interfaces, that the spaces of the screen within which looking happens start off as empty, and empty in a particular way.

An important clue that this is all more complicated than it may at first appear is the frequent and explicit conflation in descriptions of the GUI of the attitude of seeing and the relation of knowing or understanding. Norman’s praise of the “visual aids” of the GUI is one example of this. Another is a distinction made by Bruce Tognazzini between (merely) “graphical” and “visible” interfaces:

A visible interface is a complete environment in which users can work comfortably, always aware of where they are, where they are going, and what objects are available to them along the way. To be labeled a graphical interface, an interface need only make use of objects that have a distinct graphical representation. Many aspects of the graphical interface may remain invisible.

The “visible” interface is a name for the ideal to which the GUI plainly aspires: it hides nothing that would be of interest or value to the user; nothing is missing; nothing is obscured; nothing gets in the way. The much-touted usability and “intuitiveness” of GUIs depend on this myth of perceptual and conceptual transparency.

If that transparency is impaired in any way, the interface will fail Tognazzini’s benchmark: “When we set about to fool the senses through a very carefully constructed reality,” he writes, “it becomes very important that we have no hidden rules that violate the user’s sense of trust.”

In this context, the user’s mastery of objects on the screen (the formal term is “direct manipulation” or “direct engagement”) is strictly determined by the GUI’s substitution of visuality for other orders of relation. In 1982, David Canfield Smith described the substitution as the signal achievement of the new “desktop” interface of the Xerox STAR, the first commercial implementation of a GUI:

A subtle thing happens when everything is visible: the display becomes reality. The user model becomes identical with what is on the screen. Objects can be understood purely in terms of their visible characteristics. Actions can be understood in terms of their effects on the screen. This lets users conduct experiments to test, verify, and expand their understanding – the essence of experimental science.

Smith’s enthusiasm for visual catachresis (the iconic, figurative ambiguities of the interface collapse into a way of simply naming what is seen on the screen) is echoed in Hutchins, Holland, and Norman’s still broader claims for the transformative experience of direct manipulation:

The point is that when an interface presents a world of action rather than a language of description, manipulating a representation can have the same effects and the same feel as manipulating the thing being represented. The members of the audience of a well-staged play will fully suspend their beliefs that the players are actors and become directly engaged in the content of the drama. In a similar way, the user of a well designed model world interface can willfully suspend belief that the objects depicted are artifacts of some program and can thereby directly engage the world of the objects.

This alchemical metamorphosis from a “language of description” to “a world of action” is not effected simply by a technical shift from a command-line interface to a GUI (as a common misreading of Norman’s “you can just look” might suggest). A command-line interface easily can be, as Douglas Engelbart’s NLS demonstrated as early as 1968, constructed on design principles similar to those voiced by Tognazzini, Hutchins, Holland, and Norman. The thread of conceptual continuity across all these interface strategies, and the basis of any claim that an interface approaches the idealized encounters of direct manipulation, is the consistency with which they address the fields in which looking, naming, and doing take place.
Thick and Thin Spaces

Personal computing has undergone innumerable changes in the last 20 years, but in this area, most of the effort seems to have gone into technical refinement rather than critical investigation of reigning scopic and spatial conventions. Aqua, the interface of Apple’s recently-released OS X, may be the most programmatic encoding to date of these conventions.46

Hard-core fans of the command line are likely to dismiss the extravagance and graphical nuances of the Mac OS X desktop as so much eye candy, a constrained computing environment masquerading as an interactive space by virtue of putting on an especially showy dress.47 But this complaint misses the real aim of Mac OS X’s lush visual redesign, which is to bind cultural conventions of spatial complexity, depth, and transparency to practices of computing that don’t fully conform to those conventions. The masquerade of visual depth is, in an important sense, precisely so much eye candy, a constrained computing environment masquerading as an interactive space.

The history of the GUI suggests that overly-rigid implementations of screen-based interaction in “real-world” forms are destined to fail. Microsoft’s Bob interface for desktop computers and General Magic’s Magic Cap interface for handheld communicators are good examples of this fatal strategy. Mullet and Sano argue that these schemes must fail, not because they aren’t “real” enough to fool the eye, but because their crude literalism works against the need for some kinds of digital data to be manipulated in ways not tied to visual depth.48 The fictions of the “real-world” interactive spaces, for example, fracture as soon as users “open” their check registers or address books, where they are confronted by visual fields that (at best) relegate drop shadows, translucency, and the like to the margins. Successful GUls, therefore, tend to apply strategies of frank spatial hybridity, mixing flat and deep visual fields. For example, a drop-down menu will cast a subtle shadow on the objects “behind” it, but the menu items are displayed on the plainest of fields, and the letterforms will have no dimensionality.49 These hybrid approaches do not, however, challenge the basic spatial production that acts as the conceptual and procedural support for the rest of the interface. Whatever inconsistencies appear in the visual framework of the desktop remain subject to an overarching representational logic that fuses spatial depth (more precisely, spatial emptiness) and the user’s efficient manipulation of the desktop.50

Visual conventions of spatial depth and manipulation have played a more conspicuous role in the evolution of video gaming during the period of the GUI’s rise to dominance.51 Consider a design trajectory, figured on one end by the screens of ColecoVision’s 1981 release of “Donkey Kong”; one of the first video games to move beyond the purely planar schema of early games, like “Pong”; “Space Invaders”, or “Pac-Man”; and, on the other, by the more complex and finely rendered spaces of the 1999 release of “Tomb Raider III: The Adventures of Lara Croft”. Lara runs, tumbles, and blasts away at her opponents within spaces more elaborate and subtle than those inhabited by Mario and Donkey Kong.52 The responsiveness of objects depict-
ed on the screen to the user’s keyboard, gamepad, or joystick has increased enormously, in both quickness and variety. But the fundamental spatial tropology (that of space: abstract space, empty space, space that doesn’t get in the way) is consistent from the earliest to the most recent examples of game play.

As is true of most GUIs, the visual fields of video games are typically hybridized in certain ways. The game’s action ostensibly takes place in two- or three-dimensional domains in which objects, people, monsters, etc., look pretty much as they might in a “real” world (that is, they aren’t emblazoned with titlebars, menus, buttons, and the like). But game designers, facing the need to communicate vital information that can’t be gleaned from action on the screen (“How many lives does a character have remaining? How many bullets are in her gun? What’s the current score?”), resort to the use of counters or controls displayed over the game play, in the margins, or called up with a special keystroke. These visual inconsistencies don’t programmatically challenge the overall fiction of spatial openness. Their usual position in the foreground or periphery of the game window or screen reinforces the illusion that the events of gameplay somehow take place behind them. They are almost always a minor element in the game’s graphic design, which is overwhelmingly dimensional, and designers will go to great lengths to give these violations of the space of the game a look and feel that evokes the game’s visual sensibilities.

The principles of direct manipulation in the modern GUI, and play in the spaces peopled by Mario and Lara, are grounded by a single conceptual commonplace: the assumption of a prior permeable field of agency, free or nearly free of resistance, or marked only by the sorts of resistance that a more efficient keypad, a faster processor or video card, or a more “intuitive” or “natural” visual metaphor, might eliminate. I’ve been referring to this field of agency as “empty” or “transparent” space (its idealized instance), but a more accurate term would be “thin” space—a form of space that is very nearly emptied out beforehand, so that movement within it and mastery of the objects it contains are minimally challenging to users. In an important sense, users are constituted as users by their successful penetration into and traversal of this space. This is what the startup screen of the GUI signals. The desktop icon zooming out into a directory window; the expressly cinematic full-motion video sequence that “sets up” the story of the game; the constant running, tumbling, flying down corridors, tunnels, and narrow alleyways—the first effect of the graphic interface is to open up a space before you, already thinned out, ready for your purposive movement inside.

This space will take on different casts depending on the contexts of its production. It will be shaped and bounded by requirements of input devices, screen sizes, rendering speeds, and OS conventions and fashions. But its underlying structure is consistent and decisive. Putting “knowledge on the screen” (Norman really means “behind” the screen—the distinction is not inconsequential) is possible (conceivable) only if the shared domain of users’ eyes and the objects they observe may be freely traversed by them or the avatars who act in their place. Norman misses something vital to understanding the spatial regimes of the human-computer interface when he observes that there was “nothing” in the black void of the C-prompt, until the GUI revealed its secrets. There was—there is—a very particular sort of nothing, a nothing that prepared the way for the expectation that something may come to be in its place.

The Spatial Regimes of Human-Computer Interaction

The real world is the best user interface there is. And it’s an invisible interface. Or at least it’s something we’ve all learned. So we tried to make something that was as close to the real world as possible, and that meant the absence of any kind of computer interface, like buttons and things like that.

— Robyn Miller, The Making of Myst

Robyn Miller’s claim that the “real world” is an “invisible” interface is not a contradiction of Tognazzini’s praise of the “visible” interface. Both are versions of an epistemic scheme that also undergirds the exuberant rhetoric of direct manipulation: the ideal interface would be the thinnest of interfaces, the interface in which manipulation is direct manipulation because its field has been conceptually and procedurally emptied out before the interaction begins. This form of space is not, however, a given condition of interaction. It is produced and sustained by historically and culturally bracketed understandings of visibility and spatial form. Discursive practices of spatial emptying are among the most privileged methods of conceptual and political coercion of the post-Enlightenment period. They are also, as I have noted elsewhere, among the methods by which specifically cybercultural regimes of spatiality pattern themselves on permicious traditions of scientistic positivism, national-political identity, and social normativity.

An important step toward a critical-theoretical understanding of the peculiar spaces of human-computer interaction lies, I suggest, in learning to think carefully about the forms of space presupposed by the GUI and the fields of game play. These are not uncontested domains. As Martin Jay has observed of practices of visibility of the modern era, it is most accurate to say that there are multiple, overlapping, and inconsistent scopic regimes at work in the art and science of our time. This is true as well, I think, of the spatial regimes of the GUI. The history of contested spatiality in art, science, and politics off the computer screen can point the way toward a careful spatial design of the computer screen. Those debates may even provide strategies of design that break the epistemic confines of direct manipulation. The visual fields of contemporary GUIs are irreducibly hybrid; their insecurities demonstrate technical and conceptual limits of the common instances of this odd sort of place we call “cyberspace.” Every space—and this includes the spaces of the human-computer interaction—will be at least a little thick: impermeable,
imperfectly or incompletely manipulable, stuck in historical, cultural, and psychic materiality that stops up efforts to empty it out. We need better ways of looking, where the space of looking and knowing thickens.

Endnotes
n1. That the icon- and window-laden fields of GUI screens are also called “desktops” demonstrates the remarkable efficacy of the GUI. As Ted Nelson pointed out more than a decade ago [16], these images don’t look at all like the surfaces of desks. But the effect of a strong metaphor is such that it tends to eliminate from our awareness experience that doesn’t fit that metaphor.

n2. Tognazzini’s use of the term “visible” is evocative of Norman’s use of that word in the design of everyday things: “‘The user needs help. Just the right things have to be visible: to indicate what parts operate and how, to indicate how the user is to interact with the device. Visibility indicates the mapping between intended actions and actual operations’” [17].

n3. Neal Stephenson’s criticism [22] of the pervasiveness of the GUI and the decline of the command-line interface — that the former oversimplifies what the latter reveals to be complicated — differs from Norman and Tognazzini’s celebration of the GUI only in Stephenson’s preference for text and syntagm over window and menu. All three critics begin with the assumption that human-computer interaction may be (or should be) exhaustively encoded in forms of the screen, that the secrets of the system’s inner domains may be revealed to the user who is able to interpret the appropriate glyphs.

n4. For discussion in a similar vein, see [11].

n5. OS X is not unique among GUIs in using translucent widgets and antialiased shadows to suggest visual depth on the monitor screen — though Apple’s new GUI may be the most complete and consistent implementation of these schemes. Recent releases of the K Desktop Environment for Linux (KDE), for example, have incorporated some translucent elements. Windows XP, Microsoft’s recently-announced GUI for future versions of the Windows OS, also appears to incorporate similar elements. Given Apple’s traditional role as a pathbreaker in the personal computer industry, it is probable that other OSes will adopt traits of OS X, and widget translucency is likely to be among them. On the use of transparent and translucent interface widgets in general, see [1].

n6. This is, in a somewhat caricatured form, Stephenson’s complaint [22].

n7. Contrary to often-repeated claims that early cinema goes naively confused images on the screen with “real” objects and events, this was clearly not the case [5].

n8. See Mullette and Sano [15]. On the larger historical question of the role of perspectival technique in Western spatial thought, Dimosch and Panofsky [2, 18] are valuable sources. But see also Elkins [13] for a more subtle imbrication of spatial thinking and artistic practice. He argues (convincingly) that the Renaissance inventors of linear perspective never mistook it for a unifying optical practice (as Enlightenment critics would have it, and as perspective is now widely understood), but accepted it as only one of the tools available to the painter. Jay [10] emphasizes that the varieties of scopic and spatial technique in Western art and science have been far more varied and inconsistent than claimed by historians of what he terms “Cartesian perspectivism.”

n9. The use of anti-aliased screen fonts in menus and window titles aims at improving their readability, not creating an illusion of depth.

n10. For the purposes of simplifying this (very schematic) historical overview, I won’t distinguish between spatial discourse specific to coin-operated arcade, television console, and desktop computer games. I follow Poole’s lead in labelling all of these forms, “videogames” [19].

n11. See Le Diberder and Le Diberder, and Wolf [22, 24] for differing taxonomies of spatial representations in these early videogames. See also Poole’s discussion [19] of the early history of videogameing, and the evolution of three-dimensional gameplay.

n12. Though that distinction may not be true of “SuperMario 64” (1996), which discarded the platform architecture of the classic Mario games in favor of more complex spatial representations — in many ways, resembling those of “serious” action games, like the Tomb Raider series. Many Mario purists have complained that the newer, more spatially “realistic” variation of the game has sacrificed much of the charm and conceptual simplicity of the original.

n13. The cognitive significance of these interruptions in the game’s visual orders — they are signals to the user that the events on the screen are embedded in a larger psychic and cultural dynamic — has been, I think, underestimated. See [7].