Situated Mappings: Augmented-Reality Clay and Adaptive Interfacing

by

Julia Litman-Cleper
B.A. Arts Studio
U.C. Davis, 2012

SUBMITTED TO THE DEPARTMENT OF ARCHITECTURE
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE IN ARCHITECTURE STUDIES

AT THE

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

JUNE 2016

Signature of Author: Signature redacted

Design and Computation
Department of Architecture
May 19, 2016

Certified by: Signature redacted

Terry Knight
Professor of Design and Computation
Department of Architecture
Thesis Advisor

Certified by: Signature redacted

Ruth Rosenholtz
Principal Research Scientist
Vision & Graphics, CSAIL
Thesis Advisor

Accepted by: Signature redacted

Takehiko Nagakura
Chair of the Committee on Graduate Students
Department of Architecture

Situated Mappings: Augmented-Reality Clay and Adaptive Interfacing

by

Julia Litman-Cleper

Submitted to the Department of Architecture on
May 20th, 2016 in Partial Fulfillment of the Requirements
for the Degree of Masters of Science in Architecture Studies

Abstract

Our collectively developed methods of structuring information through graphical and natural user interfaces (GUI, NUI) largely emphasize visual and visuo-spatial representation over other types of sensory information. As our interfaces continue to develop we seem to select for fidelity of visual stimulation, while neglecting the behavioral aspects of physical materiality. In this thesis, I advocate for the use of expressive mediums of material engagement as part of the design of interaction within interfaces. I present an approach to interfacing using Computer-Mediated Material Interaction (CMMI), by augmenting the visual appearance and behavior of clay. This approach considers the subject's situated learning as the crucial factor in building coherent and immersive interfacing. Taking precedents from the domains of perceptual science and cognitive ecology, such as information-search, affordances and adaptive learning, I suggest that the sense of immersion and ease of learning in an interaction stems from how much it supports the subject's self-guided, iterative, discovery process. To illustrate this approach, I present a series of oil clay augmented-reality prototypes that encourage situated mappings to be discovered by the subject. I also discuss how interface design could further honor the adaptive principles latent in human-material-interactions.

Thesis Supervisor: Terry Knight
Professor of Design and Computation, Department of Architecture
Thesis Supervisor: Ruth Rosenholtz
Principal Research Scientist, Vision & Graphics, CSAIL

Keywords: User Interface Design, Interaction Design, Perceptual Science, Computer-Mediation, Material Interaction, Learning, Making, Adaptation, Tool-making, Discovery

Situated Mappings: Augmented-Reality Clay and Adaptive Interfacing

by

Julia Litman-Cleper

Submitted to the Department of Architecture on
May 19th, 2016 in Partial Fulfillment of the Requirements
for the Degree of Masters of Science in Architecture Studies

Abstract

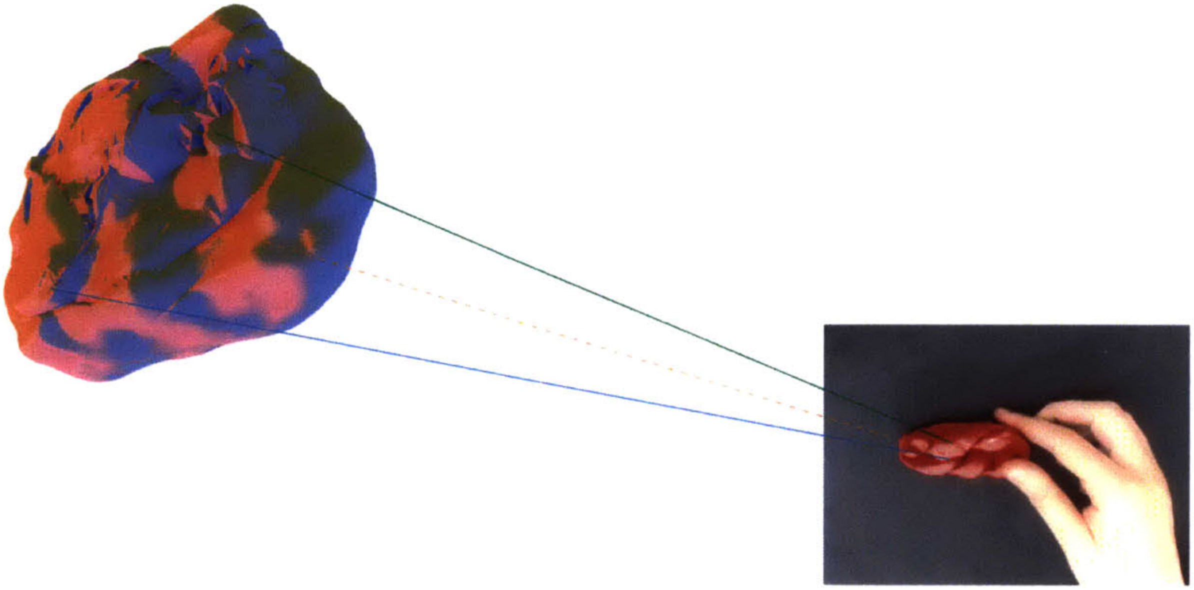
Our collectively developed methods of structuring information through graphical and natural user interfaces (GUI, NUI) largely emphasize visual and visuo-spatial representation over other types of sensory information. As our interfaces continue to develop we seem to select for fidelity of visual stimulation, while neglecting the behavioral aspects of physical materiality. In this thesis, I advocate for the use of expressive mediums of material engagement as part of the design of interaction within interfaces. I present an approach to interfacing using Computer-Mediated Material Interaction (CMMI), by augmenting the visual appearance and behavior of clay. This approach considers the subject's situated learning as the crucial factor in building coherent and immersive interfacing. Taking precedents from the domains of perceptual science and cognitive ecology, such as information-search, affordances and adaptive learning, I suggest that the sense of immersion and ease of learning in an interaction stems from how much it supports the subject's self-guided, iterative, discovery process. To illustrate this approach, I present a series of oil clay augmented-reality prototypes that encourage situated mappings to be discovered by the subject. I also discuss how interface design could further honor the adaptive principles latent in human-material-interactions.

Thesis Supervisor: Terry Knight
Professor of Design and Computation, Department of Architecture

Keywords: User Interface Design, Interaction Design, Perceptual Science,Computer-Mediation, Material Interaction, Learning, Making, Adaptation, Tool-making, Discovery

Situated Mappings

Augmented-Reality Clay and Adaptive Interfacing



For Billy

Acknowledgments

Thank you to Terry Knight and Ruth Rosenholtz for your time, thoughts, integrity, feedback, inspiration and support. Thank you to my first advisor, Joseph Dumit and to Darrin Martin and Annabeth Rosen. Many thanks also to Patrick Henry Winston, Fox Harrell and Dennis Shelden for critical and inspirational conversation.

Thank you to my colleagues, especially Cagri Hakan Zaman, Onur Yuce Gun, Diego Pinochet, Merav Gazit, Innes Ariza. Thank you to my friends Alex Cruse, Brad G., Emily Binet Royall, and Noah Kelly for the ongoing conversations. Thank you to all the student participants who lent their time and curiosity to this thesis project.

Thank you to my family, Dinetta, Mally, Hery, Aviva, Arlen, Felicia, Tonyanna, Riri and Mark. Especially to my parents for the foundational encouragement to continue pursuing questions. Thank you to The Exploratorium, and the city of San Francisco.

Table of Contents

Introduction

If you give children some clay and ask them to make spoons, they will easily know about the tools available to them, their two hands. Suppose you ask them to make spoons using a 3D modeling program, Sculptris for example, which presents a sphere of virtual clay and a series of buttons with symbols that manipulate it in different ways. Depending on that child's experiences, the process might take much longer than with the clay. It will use a different set of skills and intuitions to learn out how to control the virtual visual form, using the correlations between the virtual clay form and the mouse or device.

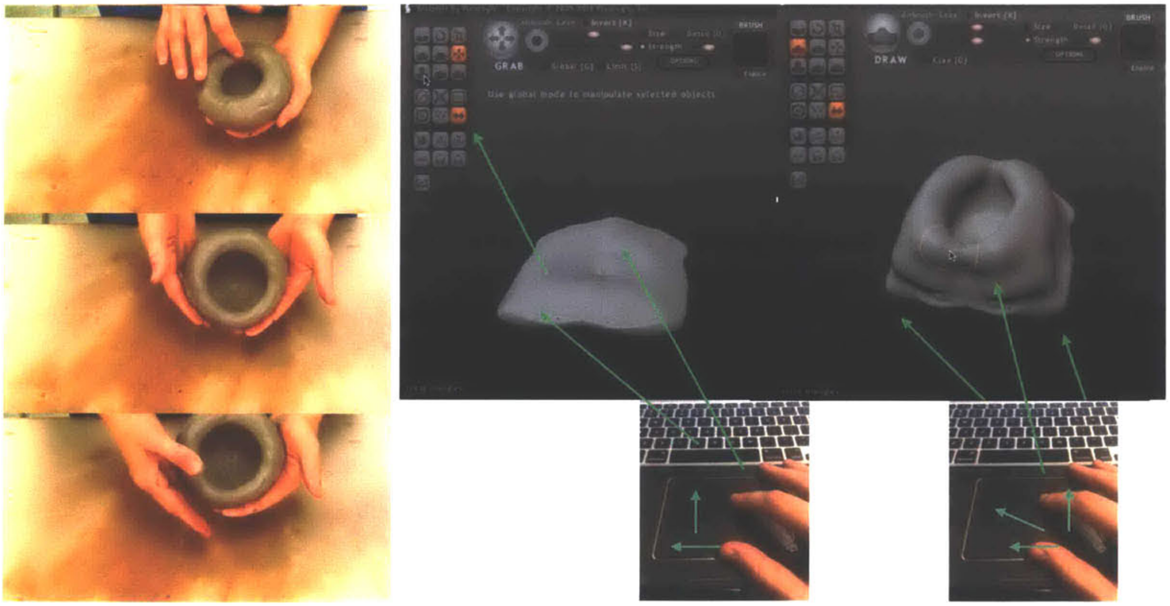


Figure 1: Material Interaction (left) versus Virtual-Material Interaction (right).

Five fingers is, so far, the maximum for all modern tetrapod hands and feet [1] and a normal human hand is said to have 27 total degrees of freedom. [2] Through this evolutionarily constrained morphology; the sophisticated measuring device that is the hand — in coordination with other sensory experiences — affords humans an ability to understand material behaviors intuitively. The strength of the learning which

---

[1] Tabin, Clifford J. "Why we have (only) five fingers per hand: Hox genes and the evolution of paired limbs." Development 116.2 (1992): 289-296.

[2] A.M.R. Agur and M.J.Lee. Grant's Atlas of Anatomy. Lippincott Williams and Wilking, 10th Edition, 1999

comes from interaction with the material world, allows humans to distinguish between materials from a single glance of an image.[3] We have an evolved ability to know the material world around us through our hands. In the process of learning a new interfaces, we map this understanding into our interfaces, and into the simulated spatial depth of our 3D modeling programs. Within these modelling programs (Figure 1, right), the visual-feedback received is "real" but the sense of spatial depth is an illusion or is "virtual." There is a mapping of the evolved set of rules, that we intuitively learn through the degrees of freedom of our own physical movement, onto an, also evolving, set of rules that are designed for a given computer program, its visual content, and its interface. I call this network of transformations, a *situated mapping.*

The ability to discern materials from visual images contributes to our understanding of affordances within other visual representations of scenes. We can look at photographs which are unreal and find rich narratives, we watch films and "willfully suspend disbelief." Yet, there is something very different about the way we use visual representation in the case of computer software, such as CAD, gaming environments, and the internet. In these cases there is the added dimension, not only of simulated spatial depth, but a "5th wall," of interactions. Guided by self-controlled motion across a given interface, we have the agency to make changes in both the visual form and the information retrieved. This is called "interactivity." Our agency, the degree to which we can change and control information, depends on the affordances and constraints of our interfaces. In standard cases this is a mouse, touch-screen or keyboard.

Situated Mapping is a broad term which signifies the context in which a person maps causes and effects to a set of affordances and constraints, building up an assumption of underlying rules. I borrow the word "situated" from Donna Harraway's "Situated Knowledges" which was written in response to the philosophical question of objective truth in the sciences. Harraway suggests, (in a simplification here) that there is no one objective truth. Rather there a plurality of objectivities which are constructed through a network of relations, situated in a context.[4] Similarly in the case of interaction-design, the complexity of the person interacting, their previous experiences, their degree of computer literacy, familiarity with a specific hardware or software and their physical ability, are all a part of the dynamics of interaction. These are symbiotically structured together, or have to be thought of as part of a given mapping.

While this implies that mappings are highly specific to situations, there are generalizable aspects of the concept as well. Two examples are the on/off button of a computer, and the Emergency stop button of an elevator, a computer-numerical-controlled Router (CNC), or other machinery. The on/off-button of a computer may be learned for a specific device. However in general, we have culturally designed a

---

[3] Sharan, Lavanya, Ruth Rosenholtz, and Edward H. Adelson. "Accuracy and speed of material categorization in real-world images." Journal of vision 14.9 (2014): 12-12.

[4] Harraway, Donna. "Situated knowledges: The science question in feminism and the privilege of partial perspective." *Feminist studies* 14.3 (1988): 575-599.

standard visual-language for the “on/off” button, which makes it learnable for many different devices. [5] Similarly, the emergency stop (E-Stop) for a CNC Router is bright red surrounded by a yellow ring. A specific individual will be trained that they should use this button when the machine is doing something wrong or dangerous. Once that individual has experienced a few occasions of error in the machine, it may become “second-nature” or instinctive for them to flail their arm out towards the E-stop when an error is happening. They have mapped this gesture from the noticing of an error, to the movements of their arm and hand to the red E-stop button. This E-stop button then is both specifically a situated mapping for the experienced machinist, as well as a symbol developed through cultural iteration, a culturally situated mapping indicating an emergency situation.

Computer-mediated reality (CM) is the concept of taking the stimuli of the real environment and modifying how it is perceived, by either subtracting or adding information, using a computer or device. For this thesis, I created an interface out of clay using computer-mediated graphics (Augmented Reality). The interface modifies the visual attributes of the clay as users play with it and see it on a computer screen. Typically, computer-mediated reality utilizes a head-mounted display, hand-held device, or projection mapping. In this case I use a computer screen and have participants focus their attention on the screen. The subjects interact by using the clay material, paired with modified graphics, as an interface to visual computer-generated forms. I hypothesize that this kind of interaction compels subjects to explore and discover unexpected associations which further user-agency, learning and creative experiences in human-computer interaction and computational making. When subjects learn through their own processes of interaction, they use their own discoveries as a tool. This points to ways we can construct unexpected interface dynamics, which encourage subject-guided interventions (agency), exploration and discovery. It is also offers insight into the adaptive processes which could be responsible for our intuitions behind tool-making at large.

In the first chapter of this thesis, I will discuss an augmented-reality interface prototype using clay and a series of experiments. These experiments show that interfaces which use computer-mediated stimuli can engage subjects attention through rules, or *mappings,* between the changes in the physical material and corresponding changes in the mediated visual graphics. A *Situated Mapping* is a subject’s assumption of rules which are used to guide interventions, navigate and learn an interface. The rules which are found are often interpreted differently for each user, even if they follow the same general idea

---

[5] Studies have shown that interface “universals” vary across cultures and that cultural preferences can determine the acceptance of certain interface designs. The cultural anthropology of interfaces is related but not the focus of this paper, however the forces at work in causing the emergence of culture-specific universals like this, is worth noting and examining further. See Evers, Vanessa, and Donald Day. "The role of culture in interface acceptance." *Human-Computer Interaction. INTERACT '97.* Springer US, 1997.

or intended outcome. In this sense the mappings discovered are *situated*, or specific to the subject and the sequences of events. Unexpected mappings can engage the subject further in a search for cause and effect. I hope to encourage the creation of more diverse ways of interfacing, and more exploration of different kinds of interface-logic. I conclude with possibilities for fully adaptive interfacing.

Background

> "…..we think of interface design as ..the points of contact between complex and perhaps incompatible systems…" [6]
>
> -Benjamin Bratton

The trajectory of user interfaces since their departure from command line interfaces (CLI) has been one of a search for increasing visuo-spatial fidelity and a search for making them seem "natural." Interfaces have moved from Command Line to Graphical User, and into the domains of "intuitive interfacing", "natural interfaces" and "reality-user-interfaces" (augmented and virtual "reality" which use input devices, such as a camera, infrared-camera, head-tracker, etc). Visual graphics which are thought of as seeming as "realistic" as possible, actually are not realistic, but tend to follow an evolving cultural logic of what "high-definition" visualization looks like. Following from the theories of neuronal recycling, and cultural transmission and iteration, there is no "more natural" or "more intuitive" interface technology but rather there is a way that an already existing sensory experience, informed by the real-world, fits into a graphical language. Although it is in part dependent on what graphics cards (GPUs) and computation can achieve at that point, it is also a part of a collectively created visual cultural language. Together, intuitions about an interface are formed by mappings between a subjective user-participant's experience of a pattern of sensory phenomena, fitted to the culturally built intuition of why an interaction with a particular set of graphics would yield a particular result. This is not dependent on the pattern of sensory phenomena being a perfect imitation of those experienced in the real-world (See Figure 2).

Instead of creating new kinds of interface logic, a change to the norm interface-logic often takes the form of customization on an individual level. Most of the symbols employed to suggest meanings in graphics are skeuomorphic; coming from already existing symbols of tools used in real material interactions. The reasons why this is has to do with many forces guiding the designs of interfaces, which

---

[6] Bratton, B. "The Convergence of Architecture and Interface Design." *interactions* (2008): 20-27.

themselves still lack a historical (and psychological) foundation, even though this practice has been ongoing. "Interaction design as a discipline does not have a coherent, historical discourse. It does not, as of yet, have all that much to say about how "interaction" was designed in ancient Rome…or among the Iroquois confederacy." (Bratton, 2008).[7]

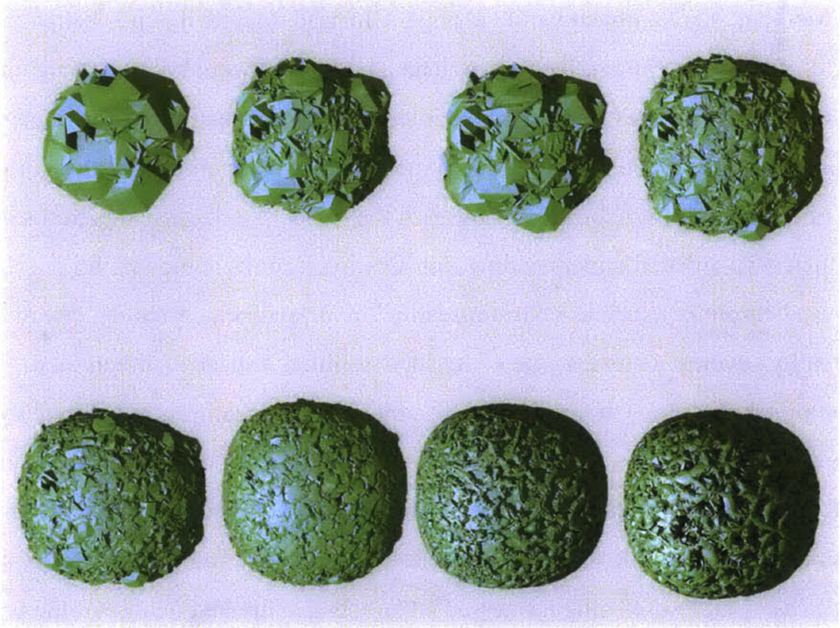


Figure 2: Although these transformations don't correlate to any "real-world" phenomena, interpretations of it as an animated graphic ("rock-like" or "root-like" are learned).

The concept of bringing interfacing into 3D space, began very early, with visionaries such as Douglas Engelbart and Ivan Sutherland. Not only was there the interest and need for three-dimensional display, but with it, an awareness that changes in correspondence with physical interactions were key. In 1968 Sutherland writes:

> "Although stereo presentation is important to the three-dimensional illusion, it is less important than the *change* that takes place in the image when the observer *moves* his head. The image presented by the three-dimensional display must change in exactly the way that the image of a real object would change for similar motions

---

[7] ibid

> of the user's head."

Acutely, Sutherland describes the logic behind head mounted displays today, which capture the the kinetic depth effect. He also expresses this notion of change, a causal correlation between head movement and an effect on the graphics of the image. I will show that the image does *not* have to change in "exactly the way that .. a real object would change." Instead, I argue that it is better for user-agency and discovery, for it to change in a *different* way than a real object would. Unexpected changes require learning and discovery. As a result of unexpected interactions, participants need to engage in active information-search. An unexpected or different mapping of a movement to visuals is more likely to allow for exploration on the part of the subject. It is through the subject's self-guided patterns of exploration that they build their own situated understanding, for how interactions could unfold.

Terms used currently, such as "virtual-reality"[8] and "augmented-reality" are misleading, because one is within "reality" even if a certain type of sensory stimuli is enhanced or deprived. Terms like "immersion" [9] and presence also can produce misleading characterizations, because an average person is technically immersed in three-dimensions for most of their life unless they have a different perceptual system. I argue that causal interaction, the rules or logic defining cause and effect within interfaces, is what captivates our attention in the process of experiencing computer-mediated sensory stimuli. These are more important to the feeling of "being immersed" than the quantity and quality of the provided sensory-stimuli and the blocking out of other stimuli. In the use of these terms one wants to refer to the definition of a threshold. This threshold indicates, through the coherency of interaction with sensory input, a structured logic within an externally designed system. Usually, the only feature which is fully illusory is the sense of spatial depth (see Figure 3) which is constructed through visual information (though it could also be constructed using sound) and through the set of causal interactions, such as rotation and scaling, that together produce the sense of a coherent simulated space.

---

[8] The word "virtual" comes from "Late Middle English....in the sense 'possessing certain virtues' from medieval Latin *virtualis*, from Latin *virtus* 'virtue', suggested by late Latin *virtuosus*." We might extrapolate this to "giving one certain abilities" in the sense of affordance, though technically this is thought to come from "in essence, potentiality, or effect, although not in form or actuality." From: "Oxford English dictionary online." *Mount Royal College Lib., Calgary* 14 (2004).

[9] Immersion refers to the degree which the user's senses are "submerged... in computer-generated stimuli...block[ing] out stimuli from the physical world" from: Biocca, Frank, and Ben Delaney. "Immersive virtual reality technology."*Communication in the age of virtual reality* (1995): 57-124.

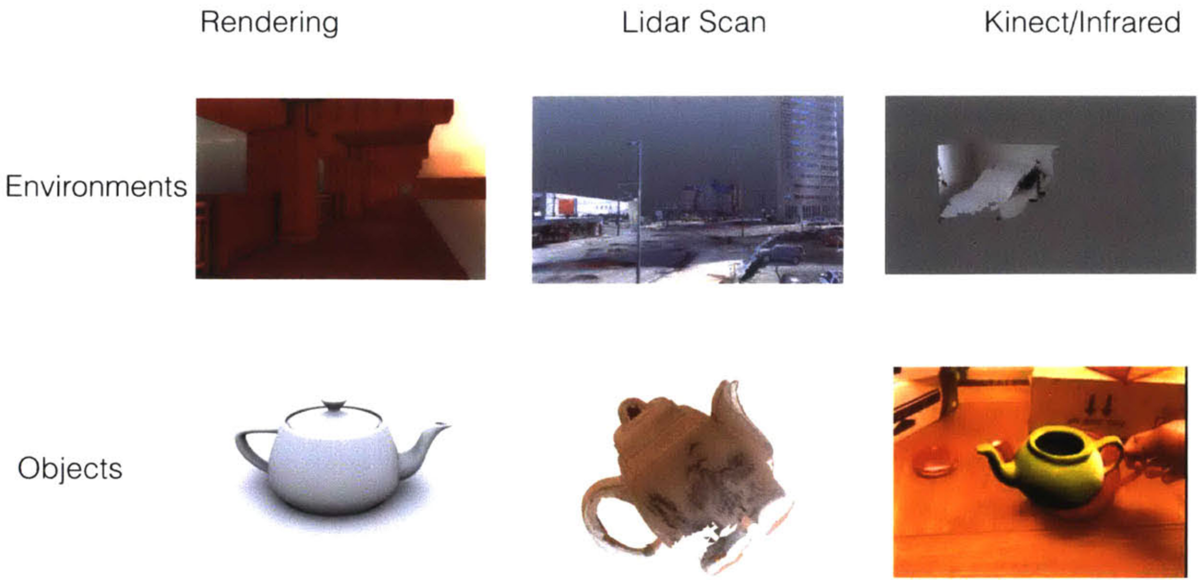


Figure 3: Types of visual environments and objects for computer-mediation. In the process of trying to achieve "realism," in 3D computer-graphics, we have ended up creating a language of uncanny, synthetic spaces that are different from our lived experiences of environments and their constituent objects. These often aren't truly "realistic," but can follow patterns of behavior that have the causal qualities of physical and spatial interaction.

Chapter 1: The Sense of "Reality"

1.1. Multimodal Perception

We often think of the perception of spaces, scenes and objects as visual phenomena, however our capacity to resolve space is possible due to the integration of multiple senses. One's sensory system is always combining different types of information. Although we think of ourselves as having 5 separate inputs of sensory information, truly separating these is difficult. Our keen ability to move through space is testament to multisensory (multimodal) integration, since spatial knowledge uses several of the senses[10] in coordination through time. As Poincare wrote:

> *"To localize an object .... means to represent to oneself the movements that would be necessary to reach it."* [11]

---

[10] Eimer, Martin. "Multisensory integration: how visual experience shapes spatial perception." *Current biology* 14.3 (2004): R115-R117.

[11] Poincaré, Henri. *The value of science: essential writings of Henri Poincaré*. Modern library, 2012. (page 47)

Even though visual stimuli makes up written language and symbolic thinking, aspects of higher-level thinking such as inference, come from a build up of experiences across multiple senses. For example, the developing child has poor eyesight, which gets better over time. [12] [13] The bulk of early stages of information intake is through the skin (tendons, muscles, fibrous capsules in the joints) and through movement (the vestibular system). The proprioceptive and kinaesthetic senses are very important to processing visual information in space. In an experiment by Held and Hein in 1963, two kittens were exposed from birth to the same visual information at the same time. [14] However, one kitten was passive by being constricted and linked the other's movements through a carousel-like apparatus. The kitten who could autonomously move processed the visual information normally, whereas the constricted one could not. This is proof that self-controlled movement through space is a key factor in the development of coherent visual cognition.

Vision is considered the dominant sense due to findings such as visual capture, where humans bias vision as true over other senses in many situations. This happens when we look at a ventriloquist or watch a movie (ventriloquist effect), in the Mcgurk Effect (vision over sound), and in the rubber-hand illusion (vision over touch).[15] However it has also been shown that vision isn't dominant always and that attention, context and language can result in the biasing of other senses over vision. In some situations sounds and tactile cues can have an effect on vision.[16] Russian speakers, who have learned special words for tones of blue can distinguish this color where others cannot.[17] Even on the neurophysiological level (across Marr's computational levels of analysis[18]) this is also a complicated boundary. For example, in monkeys it has been found that neurons that are considered to be specifically responsive to vision (visual

---

[12] Dobson, Velma, and Davida Y. Teller. "Visual acuity in human infants: a review and comparison of behavioral and electrophysiological studies." *Vision Research* 18.11 (1978): 1469-1483.

[13] Sinha, Durganand, and Pushpa Shukla. "Deprivation and development of skill for pictorial depth perception." *Journal of Cross-Cultural Psychology* 5.4 (1974): 434-450.

[14] Held, Richard, and Alan Hein. "Movement-produced stimulation in the development of visually guided behavior." *Journal of comparative and physiological psychology* 56.5 (1963): 872.

[15] Ehrsson, H. Henrik, Nicholas P. Holmes, and Richard E. Passingham. "Touching a rubber hand: feeling of body ownership is associated with activity in multisensory brain areas." *The Journal of Neuroscience* 25.45 (2005): 10564-10573.

[16] Shams, Ladan, and Robyn Kim. "Crossmodal influences on visual perception." *Physics of life reviews* 7.3 (2010): 269-284.

[17] Winawer, Jonathan, et al. "Russian blues reveal effects of language on color discrimination." *Proceedings of the National Academy of Sciences* 104.19 (2007): 7780-7785.

[18] Marr, D.; Poggio, T. (1976). "From Understanding Computation to Understanding Neural Circuitry". Artificial Intelligence Laboratory. A.I. Memo. Massachusetts Institute of Technology. AIM-357. The 3 levels of analysis are defined as computational, algorithmic/representational, and physical. This chapter suggests that across multiple levels multisensory processing is a requirement. T. Poggio. adds that a level above computational is that of learning.

receptive fields) also respond to the coordinates of the arm when the arm moves and there are visual neurons, specifically of the visual pathway, which respond to tactile stimuli.[19]

The multimodal nature of our sensory processing suggests that combinations of different types of input allow us to understand the environment. So it follows that our visually "high-definition" simulations of space, still fall short of the feeling of "real" space. What there is may be sensory adaptation to this slightly different visual-illusion, or representation of spatial depth. The term "presence" has come to signify moments where virtual environments are effective at producing the feeling of "being there" or being immersed.[20] A situated mapping occurs where sensory adaptation to cause and effect has happened coherently enough for a feeling of immersion, even when the visual stimuli may not be precisely realistic or accurate. It was found in a study on virtual reality environments that users were more likely to believe in the physicality of a virtual table in front of them if, when they moved their hand onto it as if to hit it, the sound of hitting a table was heard.[21] Even though their hand was still moving through air and not hitting any physical table, the correlation of the sound with the visual imagery of a table strongly influenced the subject's perceptions of a table, in comparison to without the sound.

These findings suggest that we do not need "realistic" visual fidelity, as often is the case in creating effective human-computer interactions in both interfaces and synthetic environments, but rather that we should leverage situations of sensory integration from correlated events. Our senses are adaptive and adjustable, they are subject to forces of attention, carved by language, context (the statistics of environments over time) and correlations among the different senses. Therefore, it isn't the nature of the sensory stimuli but the contexts and patterns of interaction over time, which shapes the percept of a "natural" interface. It is not necessary then that we simulate a material's property, for example, but rather that we create causal structures so that the interactions follow similar patterns of behaviors as to those when we experience a physical material.

In order to construct this logic, it is necessary to deconstruct those elements which produce the sense of a material's physicality. For the purposes here I separate 3 key properties that characterize a sense of the prototypical[22] material of a category: Texture (surface conditions, lighting), Shape (possible geometries) and physics behaviors (changes over time, reactions to given effectors). (See Figure 4) Each

---

[19] Graziano, Michael SA, Gregory S. Yap, and Charles G. Gross. "Coding of visual space by premotor neurons." *Science-New York, then Washington-* (1994): 1054-1054.

[20] Riva, G., F. Davide, and W. A. IJsselsteijn. "Being There: The experience of presence in mediated environments." Being there: Concepts, effects and measurement of user presence in synthetic environments 5 (2003).

[21] Conn, Coco, et al. "Virtual environments and interactivity: Windows to the future." Acm SIGGRAPH Computer Graphics. Vol. 23. No. 5. ACM, 1989.

[22] Rosch, Eleanor. "Principles of categorization." *Concepts: core readings*(1999): 189-206.

of these are also possible to augment in terms of visual stimuli. Changing one of these features too much leads to a sense of it being a different material.

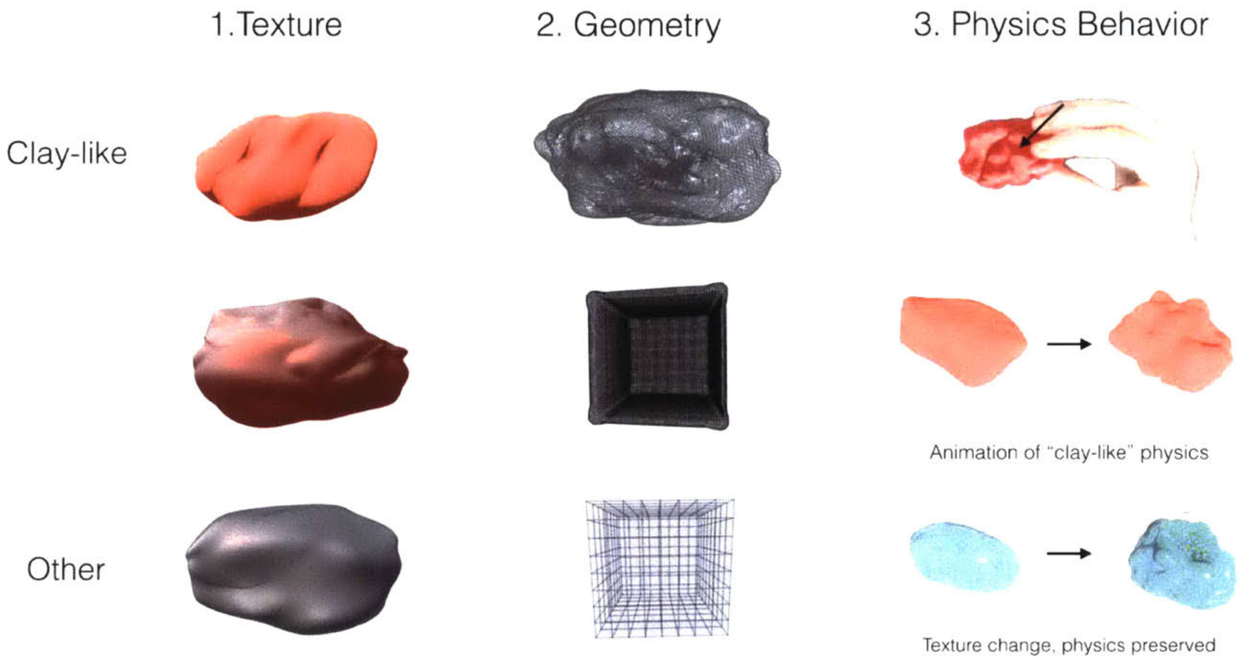


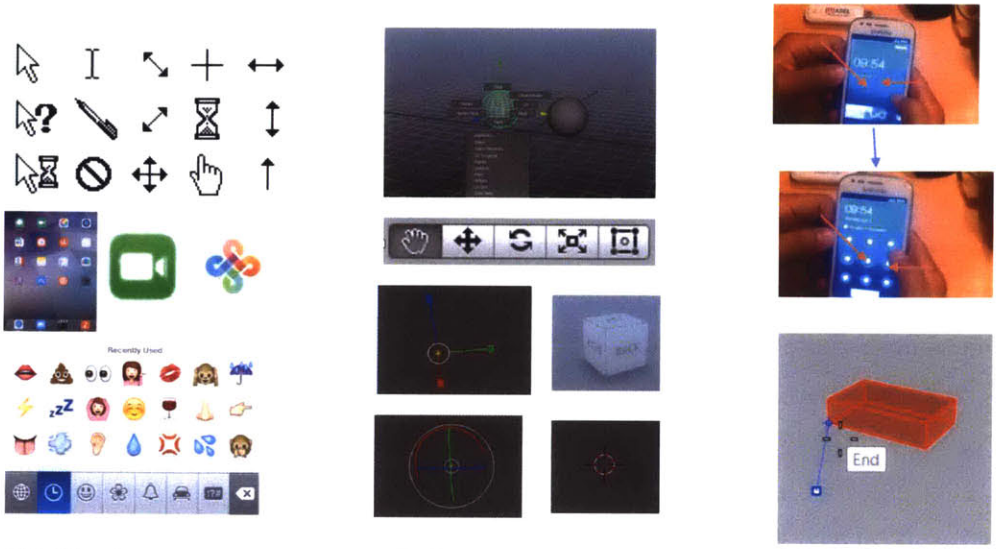
Figure 4: Material Augmentation, defining 3 key properties.

It is the correlations, the behavioral coordination, between these features (here specifically visual, though there are possibly many other features) that builds rule-structures behind material interactions. Rather than perfectly simulating the visual behavior of all real materials, I suggest that we could strive instead to create a generalized logic that is similar to the causal structures that allow us to learn, discriminate and define materials in the world to begin with.

The versatility of the senses, as well as the increasing need for information organization due to the saturation of information brought into our daily lives through the internet, makes it all the more imperative that we understand the effect of computer-mediation and interfacing on the structuring of information and on cognition. I address this need by testing how subjects learn new kinds of interface-logic and build their own idea of underlying rules. I use a combination of computer-mediation and real human-material interaction in order to create a dynamic interface. Augmented-Reality Clay serves as this interface as well as as a kind of metaphor for a raw material that could be defined as users interact with it.

1.2 Existing Extensions: Tools for Interaction

Currently many of the tools used to navigate simulated spaces and synthetic environments can be defined as either glyphs, cursors or gestures that combine these (Figure 5).



2D Glyphs 3D Cursors 4D Gestures

Figure 5: A number of types of "tools" for navigating virtual spaces. These images are taken from Iphone devices and 3D modelling software (Unity, Maya, Blender, Rhino).

These current tools might be categorized into 3 types: 2D glyphs, 3D cursors and 4D gestures. However, for 3D cursors to function they require interactions such as rotation, so are on a continuum and blend into each other. Also notable is that none of these provide actual material feedback even in the case of touch-screens. For example, "swipe-ing" on a smartphone means mapping the specific visual symbols as they correlate to gestures over time. While one receives feedback about where one's hand is spatially, and about the kind of movement required for the gesture, the actual material feedback is neglected. One is still feeling the same cover of the device without any changes. Touch is virtual on the smart-device, there is no actual haptic feedback, rather the feedback is in the visual correlation (it is hard to operate these with one's eyes closed for example). It is curious that materiality is neglected even in the case of touch screens.

These tools can all be thought of as extensions of the proprioceptive and kinesthetic feedback system in our body and they are extensions in the McLuhian sense, media extensions of communicative function. [23] The cursor, like the on-button, is a part of a historically and culturally built up language of

---

[23] McLuhan, Marshall. *Understanding media: The extensions of man*. MIT press, 1994.

interface-logic, which includes in its mapping the idea of "clicking" and the spatio-temporal aspect of "dragging" windows on a screen.

The mapping between these symbols or glyphs and their constituent synthetic spaces still isn't widely resolved, though there are many attempts and precedents for the field of 3D interfacing. The missing dimension of materiality is certainly noticed and still overlooked in 3D spatial interfacing. The nuance and degree of information gained from our natural Exploratory Procedures (EP, Figure 6)[24] is not paralleled in our interface dynamics. Devices such as the Kinect, LeapMotion and SoftKinetic all use the added hardware of infrared or depth-imaging in order to allow for spatial interaction. They typically project a grid, or rays, of infrared light and then sense where the light bounces back. This results in a picture of space, with the actual coordinates of things in the environment. Input devices like these promise to provide a smoother bridge between real movements in space and simulated visual responses in our representations of space.

Other precedents include Illuminating Clay, a project from the Tangible Interfaces group which uses a ceiling-mounted 3D laser scanner to get the dimensions of clay, and projects a computational analysis of landscape back onto the clay.[25] This is a very interesting mapping of material and simulation. However many of these interfaces require extra hardware in order to run and this hardware can be cumbersome or intrusive to the process of interaction. For this thesis project, only a webcam is needed and is hung above the table where the clay is located. By using only a webcam and computer vision one can create a much more flexible system that can be positioned above any kind of material. The idea of using "Visual sensing provides a passive and non-intrusive way for computers to acquire gesture input information. Visual modality is natural, untethered, and inexpensive." [26]

---

[24] Lederman, Susan J., and Roberta L. Klatzky. "Extracting object properties through haptic exploration." *Acta psychologica* 84.1 (1993): 29-40.

[25] Piper, Ben, Carlo Ratti, and Hiroshi Ishii. "Illuminating clay: a 3-D tangible interface for landscape analysis." *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 2002.

[26] Guan, Haiying. *Vision-based Three-dimensional Hand Posture Estimation Using Hierarchical-ISOSOM*. ProQuest, 2007.

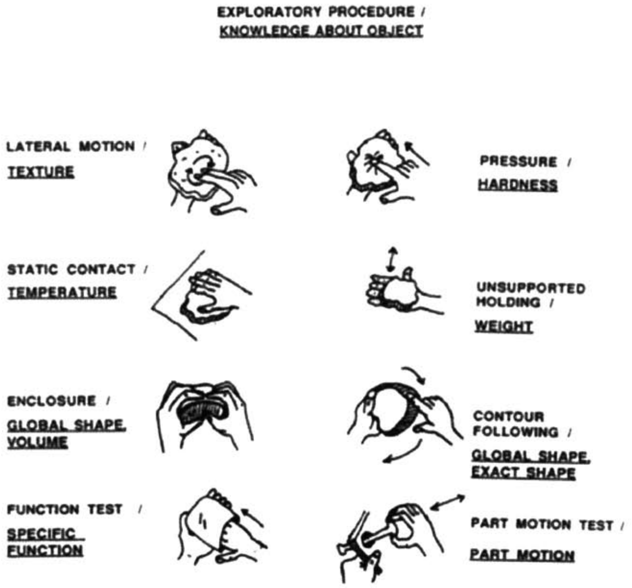


Figure 6: Exploratory Procedure, From Lederman and Klatzy, 1987.

1.3 Language and the Evolution of Interaction

While aspects of the materiality of the world are are unparalleled within interfaces and computer-mediated systems, similar human adaptive behaviors are observable as in other real-world human exploratory search processes. Spatially mediating our environment is also nothing new, as the history of architecture and urbanism can attest to the diverse development of spaces in the built-environment. But what is unique about the pairing of materials with the vast possibility space of computer-generated visual simulation is that it enables different and sustained mappings between sensory stimuli and external information. We can pair these interactions with other informational-retrieval capacities such as through the internet, which otherwise would not occur. There is much debate about how to organize information to better allow for exploratory information retrieval, in the field of human-computer-interaction. [27] [28]

Designers of search browsers for the internet, have made theoretical comparisons between types of information-search or IR, and patterns of foraging for food as an energy-time-optimization task in

---

[27] Warner, Julian. *Human information retrieval.* Cambridge, MA: MIT Press, 2010.

[28] Svenonius, Elaine. *The intellectual foundation of information organization.* MIT press, 2000.

animals. [29] One can speculate that the trajectory of the development of interaction logic, is subject to similar processes of natural selection through human cultural evolution. This is in further support of the idea that interfaces can be made unconstrained and modular enough such that they emerge out of use, through subject-driven or group-driven exploration.

In many ways speculation about the effects of interfacing are parallel to controversies about the evolution of language. The theory of cultural transmission and iterated learning,[30] suggests that human cognition is developed by our ability to learn socially as a culture, across generations. [31] So the process of developing concepts and collective design are subject to the dynamics of evolution through cultural transmission. We are able transmit complex ideas and carry out the process of iterating and refining them to a much greater degree than other communal species, due to the transmission of information across generational time scales. It is possible then that new kinds of interaction-logic, as discussed earlier, could emerge through collective processes of selection.

In the case of written language certain types of symbols have evolved over time. For example Mayan letters used face characters as letters, but processing faces may not be optimal for reading as the face area and the reading comprehension area are "far apart" in neurological structure. So it follows that these face-characters as letters would gradually become selected out or extinct. [32] Across "surviving" languages today there are many similar traits, suggesting a common visual convergence around the biologically optimal speed for detecting contrasting edges. Adaptation and evolutionary processes become circuitous. Language is selected to optimize already existing neuronal structures but it also has an effect on cortical organization and has altered certain cognitive capacities. [33]

Knowing how to use interfaces properly has become equivalent to literacy in terms of mediating access to information and to communication. Therefore, one can speculate that interaction design could produce the same forces of selection that other processes like written language have. The idea that language-change over time is a process of cultural evolution, with languages themselves being subject to evolutionary forces, is similar to other aspects of human engineered and designed systems which develop collectively over time. Interfacing is not actually "intuitive" immediately, rather, we can think of the act

---

[29] Pirolli, Peter. "An elementary social information foraging model." *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2009.

[30] Griffiths, Thomas L., and Michael L. Kalish. "Language evolution by iterated learning with Bayesian agents." Cognitive Science 31.3 (2007): 441-480.

[31] Henrich, Joseph. *The secret of our success: How culture is driving human evolution, domesticating our species, and making us smarter*. Princeton University Press, 2015.

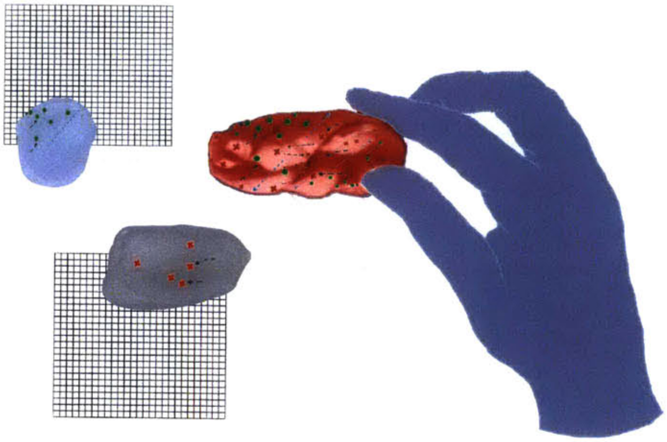
[32] Kemmerer, David. *Cognitive neuroscience of language*. Psychology Press, 2014.

[33] Dehaene, Stanislas, et al. "How learning to read changes the cortical networks for vision and language." science 330.6009 (2010): 1359-1364.

of interfacing as a kind of literacy, where the "grammatical rules" underlying interface logic are learned. These are arguably still in the process of finer optimization through cultural selection. One is faced with the same controversies in interface-logic as about language, whether or not these are grounded in some pre-existing optimal way of organizing information for our physiological condition and we are selecting for these, or whether or not we learning these systems externally and this is selecting us as it were, by slowly shaping our biological abilities. It may the push and pull of both.

Chapter 2:

Augmented Reality Clay as an Interface



2.1 Experiment Set-up

I created an Augmented Reality program that allows for any material interaction to become an input, mapping it as an interface to a simple graphics program using Processing. I use clay and 3 types of visual overlays (augmentation) to allow subjects to use clay material as a re-shapeable interface to computer-generated forms. This is done with a webcam, oil clay and a computer (running Processing, OpenCV and ToxicLibs). Since the system only requires a webcam, any material could be used, such as string, cloth or paper. I chose clay because it is very tactile, transformable and takes up a volume of space in three-dimensions allowing for the interrogation of both the physical and spatial attributes of material interaction. I chose to visually augment 2 of the features of clay: the texture and the set of perceived physics behaviors of the clay.

The main hypothesis in conducting the experiments was that even unexpected visual behavior, if it follows a coherent rule, will be controllable by users. This follows from my thesis claim that coherent situated mappings matter more than realism for interfacing.

Experiment Set-up:

1. A camera (webcam) is mounted above a black piece of paper, which is the sculpture area.
2. The camera is plugged into a computer which runs the computer-mediated visuals.

3. A monitor connected to the computer shows streaming video from the camera and the mediated visuals to the subject as they interact with the material in the sculpture area.

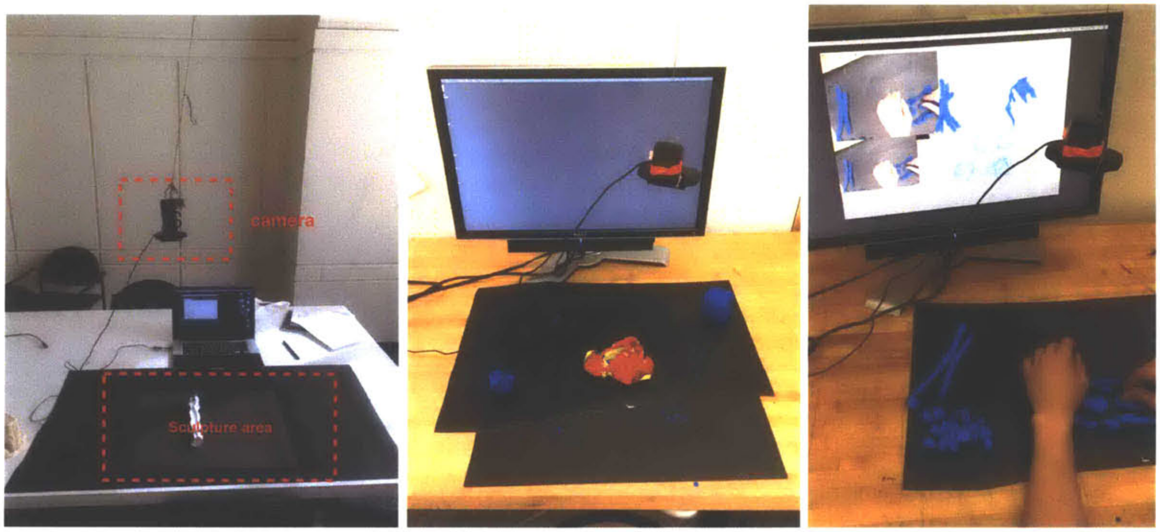


Figure 7:_Experiment setup, Two versions, set-up for pilot tests(left), main set-up (middle), main set-up in action (right).

The screen for each version of the augmentation is split into 4 (see Figure 8). The bottom left shows the live-streaming of the camera (Video Stream) which is positioned above the sculpture area where users can play with the clay. The top left shows the video stream with the augmented visuals overlayed on top (Augmented Video Stream), the top right shows the same over-layed augmented visuals on a white background (Augmentation Stream) and the bottom right shows a pre-made target drawing made with the interface (Target Drawing). Altering the shape of the real-clay (seen in bottom left of Figure 8) is synchronized to alterations in the shape of the virtual, visual forms on the screen (seen in the top left and top right of Figure 8). The augmented visuals can also be captured in time through a mouse click to allow one to build up more intricate shapes in the top-right screen (Augmentation Stream). The captured visuals remain in the top right section, this area is the equivalent of the canvas of the drawing program, which the Augmented clay acts as an interface to.

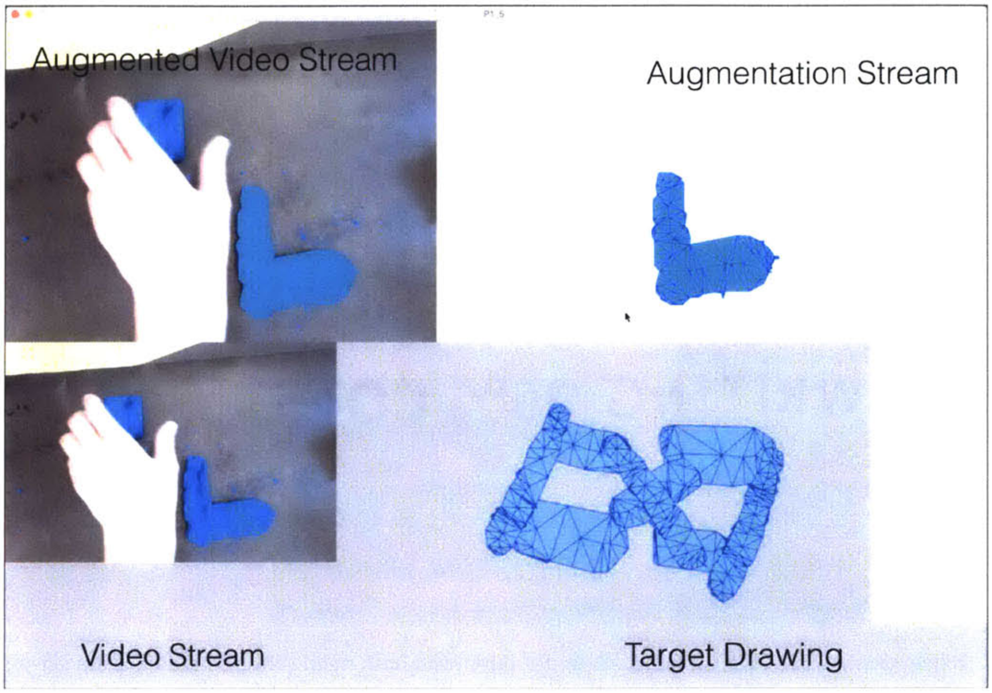


Figure 8: Screen capture (version 1, test condition 2) of the general layout.

I used 10 subjects. An additional 5 others were tested only on the third version of the overlays (animated-inflation). All the subjects were MIT graduate students except one of the additional set of 5. They were from Architecture, Comparative Media Studies, Art Culture Technology and Urban Planning.

The first version was the flat polygon visual overlay. This did not have any physics-like behavioral rules, its rules were that it overlayed the largest contour and matched the shape of the contour. It can also be recorded in the top-right canvas area. I ran the flat polygon (planar mesh) program in two conditions. For the first condition the subjects were asked to recreate a the layout of the target image, which contained squares at different sizes. I covered up the top-left quadrant, (Augmented Video Stream area) so that they only had the overlay's in coordination with their gestures in order to understand how the system was behaving and rules there were underneath this behavior. I did this so that I could compare whether seeing the clay with the augmentation helped for understanding the rules or not. Results on this inconclusive, but mainly subjects said that it did not change much.

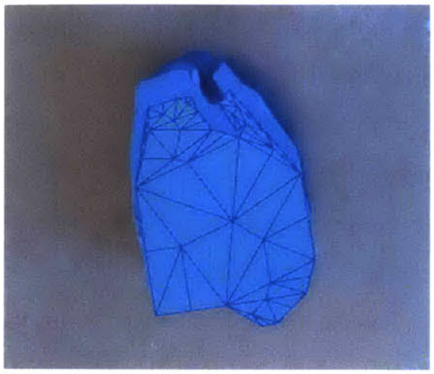


Figure 9: Close-up of clay over-layed with a triangulated flat polygon (planar mesh).

In the second test the subjects were allowed to see the full screen. This time I chose the figure-8 target image (see Figure 10). I deliberately chose this difficult-to-produce shape in order to point the subjects towards situations in which they might encounter errors, the idea was to observe how subjects would interpret certain behaviors and how they would work around it.

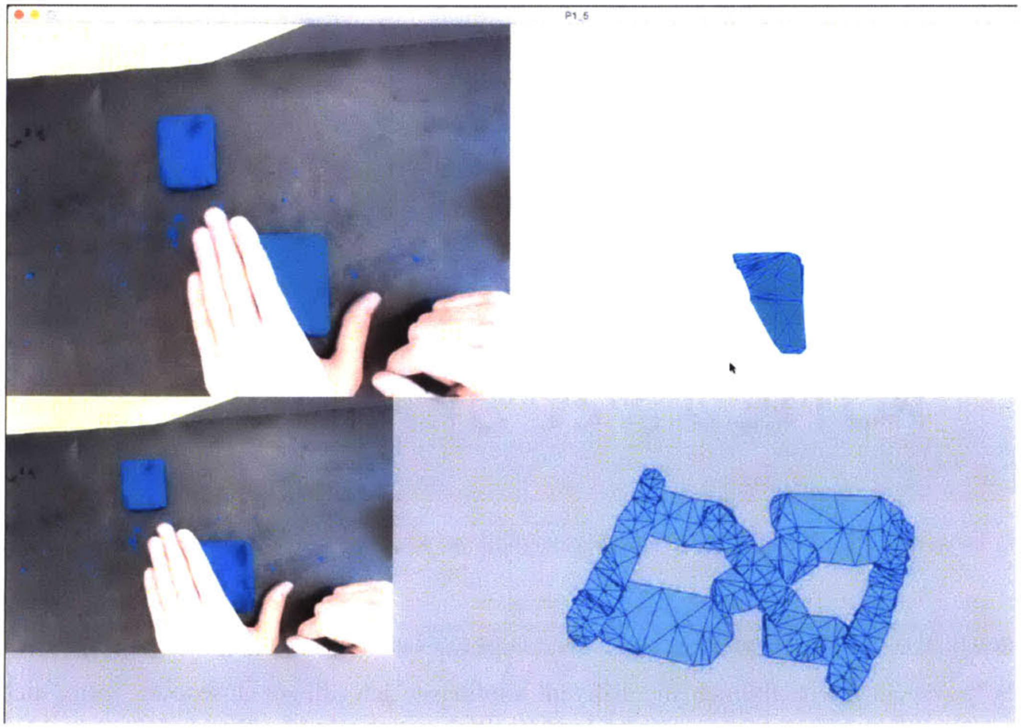


Figure 10: Screen Capture, Version 1. The contour is triangulated into a flat mesh. The shape is drawn and can be recorded in the top-right screen. The Figure-8 target condition was the second test case.

The second program takes the contour and represents it in a series of 3D spheres (voxels). The behavioral rules are that the spheres will get condensed and dispersed according to the size of the contour. The cluster of spherical voxels can also be recorded in the top right screen (see Figure 11). Similarly in the third test, I chose a fairly difficult-to-find distribution of voxels, to cause the subjects to need to explore the rules underlying the program thoroughly. However, the general shape of the target image could still be achieved in multiple ways (some of which ended up being ways I could not have anticipated), so there were no "correct" or single solutions. This created a situation of second-guessing for the participants, which encouraged the process of information search. Subjects would keep trying other things because their results were often close but not quite perfectly on, so the assumption was that there could be a better or easier way of reaching the target image.

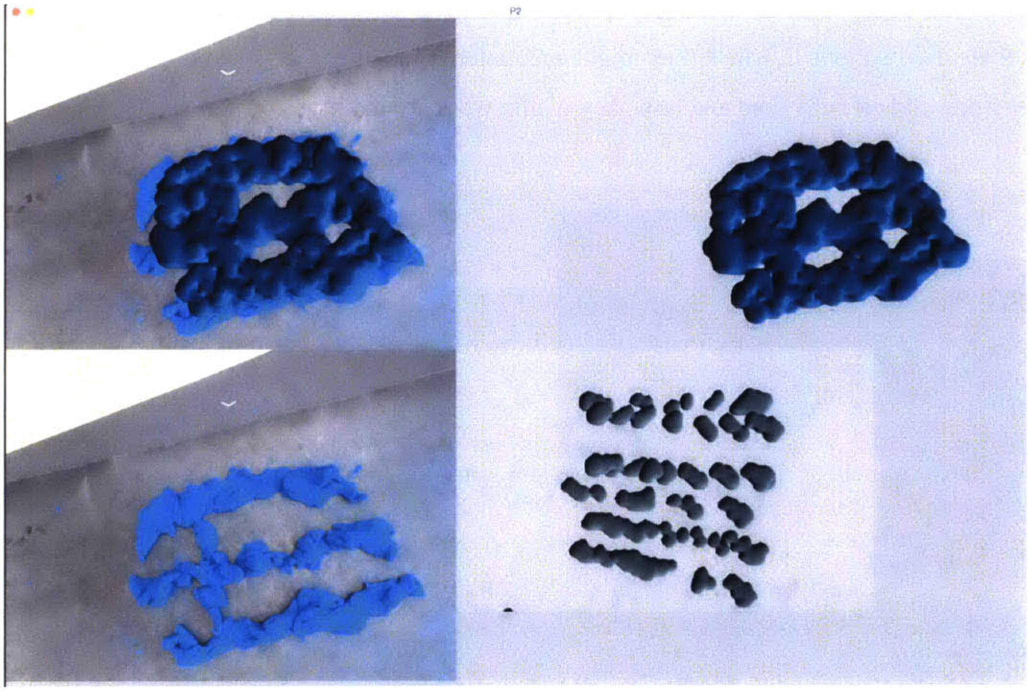


Figure 11: Screen Capture, Version 2. This was the third test case.

Figure 12 shows how the target image for the third test case was originally made using snake-shapes, however subjects found many different solutions which still got close to the form, such as a large cluster in a square (Figure 13), a grid (Figure 14) and the use of clay of another color to remove voxels (Figure 15) .

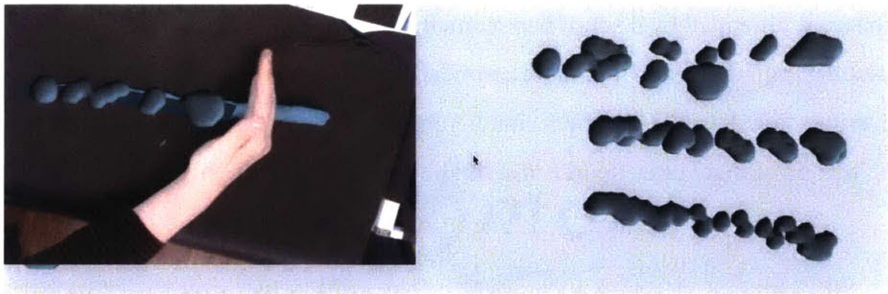


Figure 12: Original making of target image using "snake-shape."

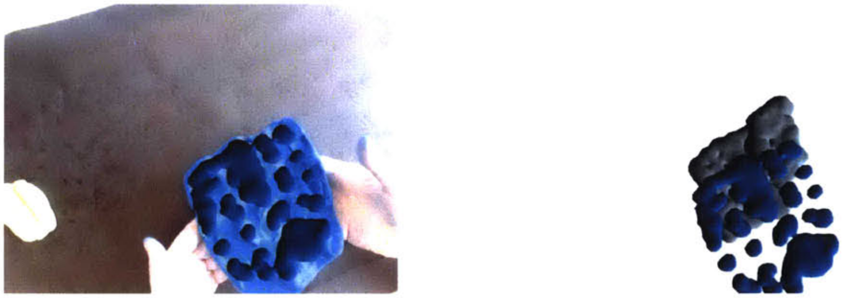


Figure 13: Solution by making a square.

Figure 14: Solution by making a grid of snake-shapes.

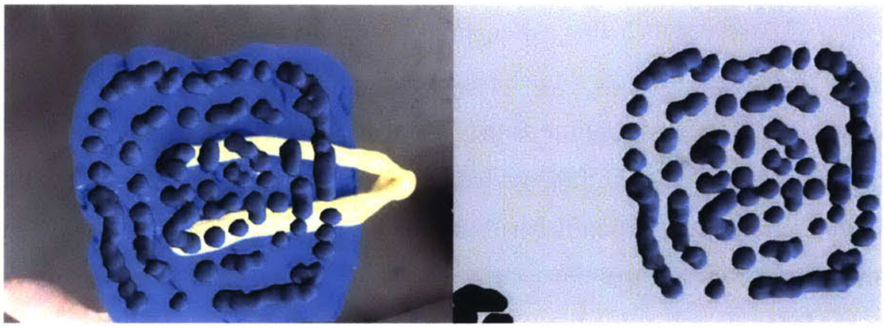


Figure 15: Solution by making a square and adding a different color to remove voxels.

The third version of the program, tested on 5 additional subjects, takes the contour and maps the size of the contour to a single sphere. The size of the sphere is relatively equal to the size of the detected contour. The sphere is animated by a set of behaviors that are modelled based on the dynamics of inflation, but fluctuating more often and more erratically (see Figure 16). This was done using a library for physics behaviors called ToxicLibs, by Karsten Schmidt. The code for this project as well as the libraries used in this project are open source and can be found on github (https://github.com/jjuliia).

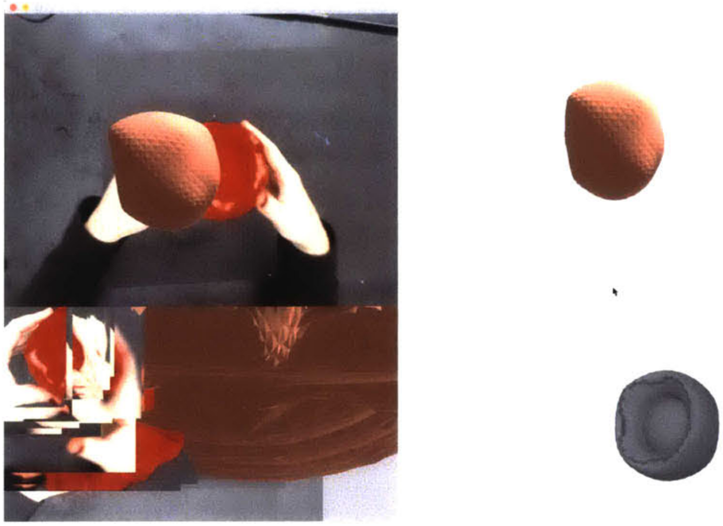


Figure 16: Screen Capture, Version 3

Participants in the experiments were asked to sit down in front of the clay and attend to both a large monitor and to the area where the clay was. They were told that the clay corresponds to imagery on the screen in some way. I asked them to first get comfortable with the clay, and think about if there are any overall rules or a sense for how the clay may be corresponding to the imagery, and that when they are ready to let me know. I observed their first assumptions and gestures in guessing about the system. Once comfortable I asked them to try to make something similar to the target image, in the bottom right hand screen, and that they could use as much time as they wanted to. Participants were also allowed to opt-out of any of the versions of the program if they wanted, this happened mostly for users who didn't want to spend too much time for all three. I also informed them that it was possible to record moments in the top right screen with a mouse click and that they should tell me if there are moments when they would like to freeze an image. I tried to reveal as little as possible otherwise.

For all procedures I encouraged any narration and noted their main points. I also asked 3 questions after they had finished. The first was, which version they found the most difficult to learn. The second was, what rules and behaviors seemed to be at work in what they noticed. The third was, which screen they found themselves focusing on most during interacting and why. I also took video recordings of the screen for each version of the program, and informed participants of this recording would happen, and would be saved but kept confidential.

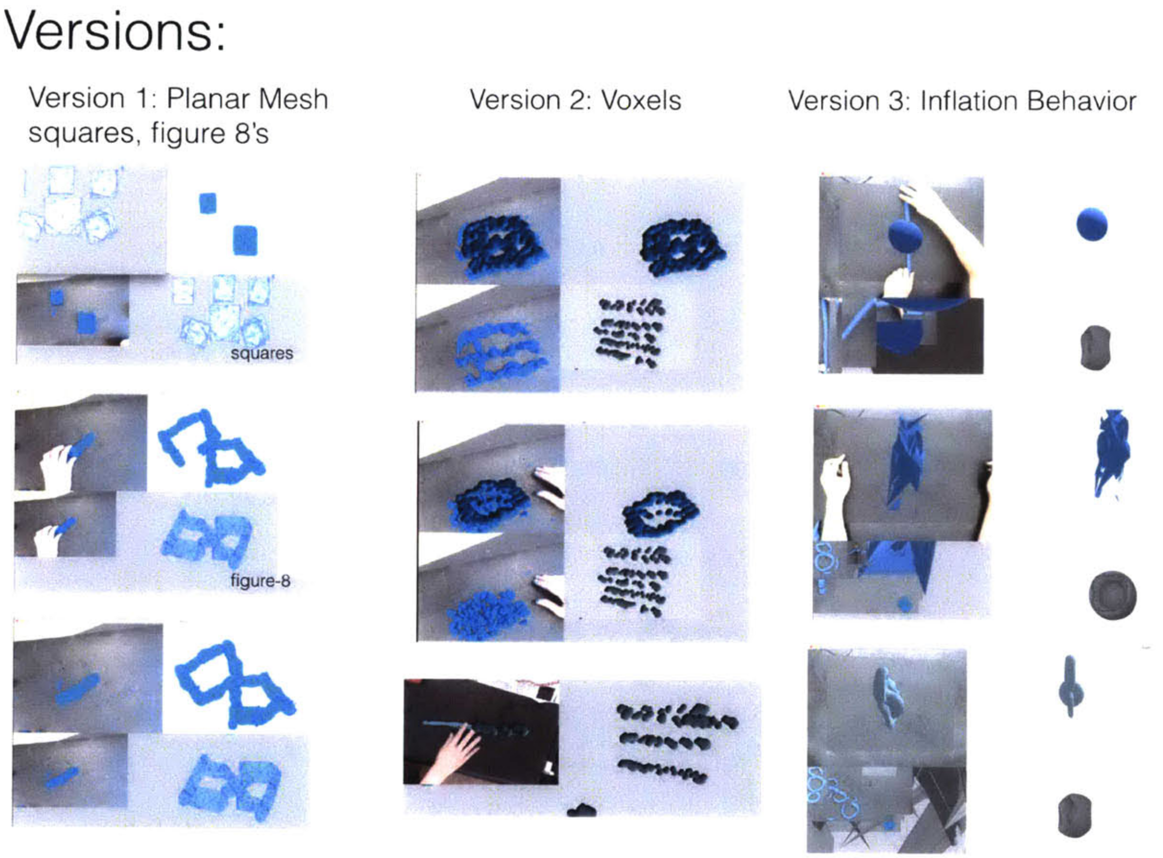
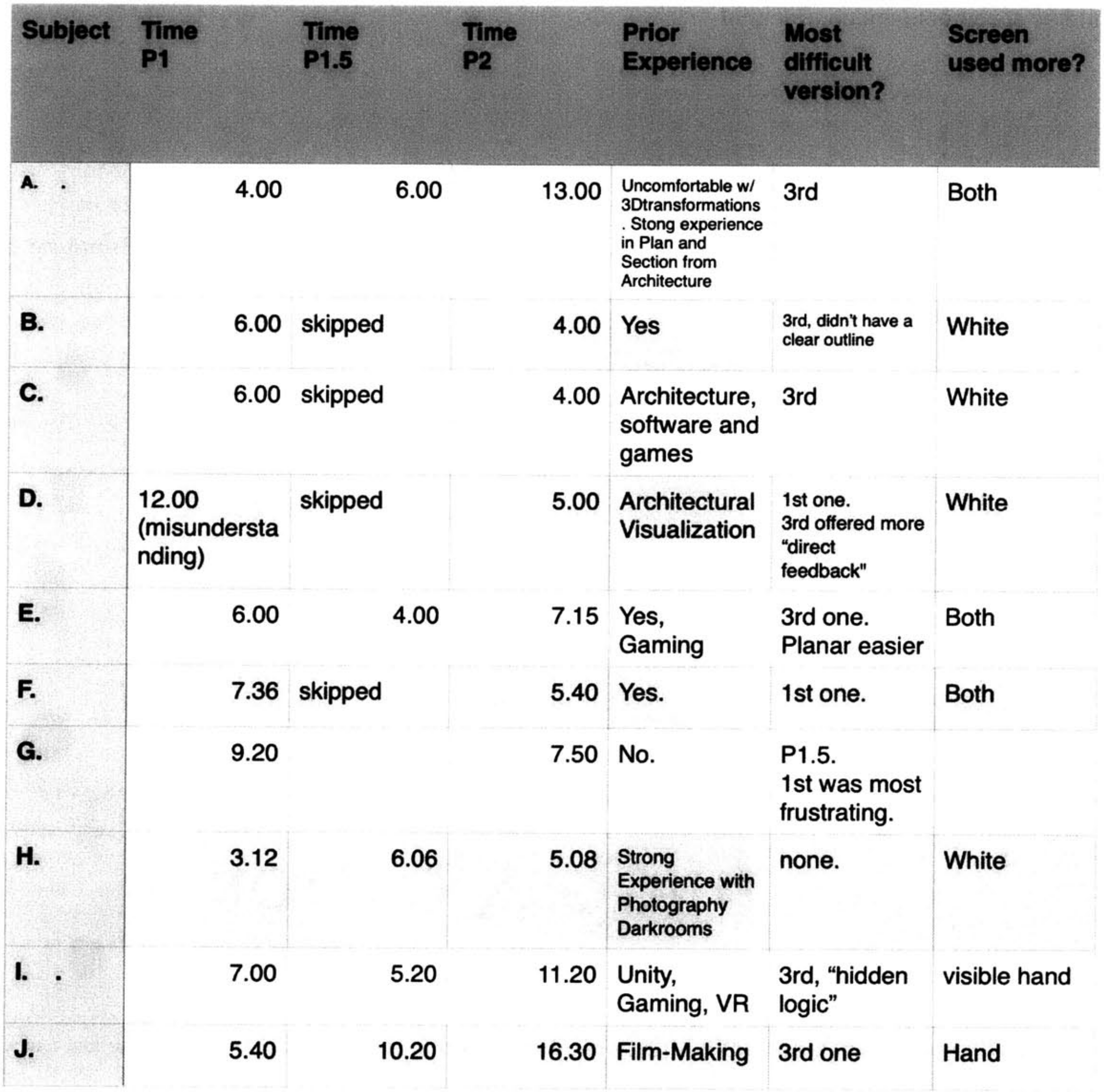


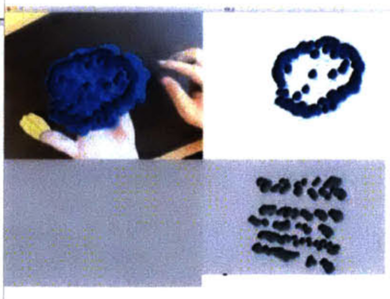
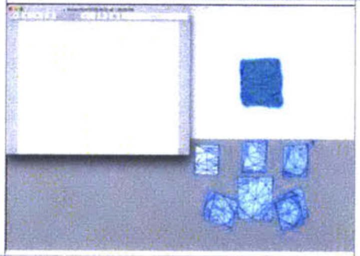
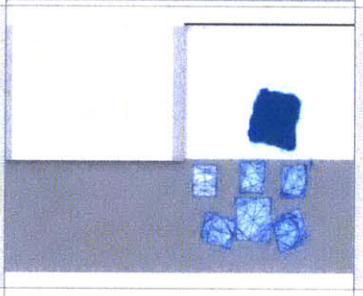
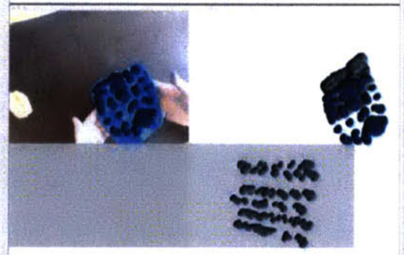
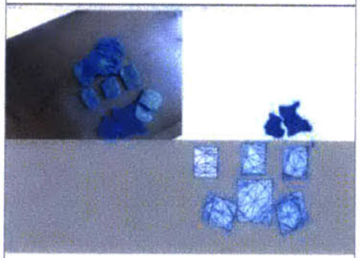
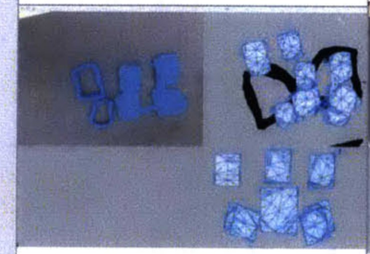
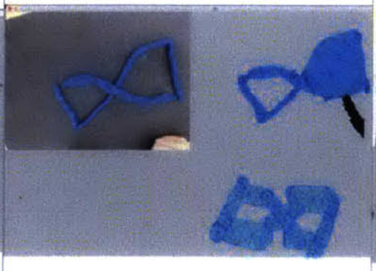
Figure 17: A side by side comparison of the 3 Versions of visual overlay (augmentations) used for the test cases.

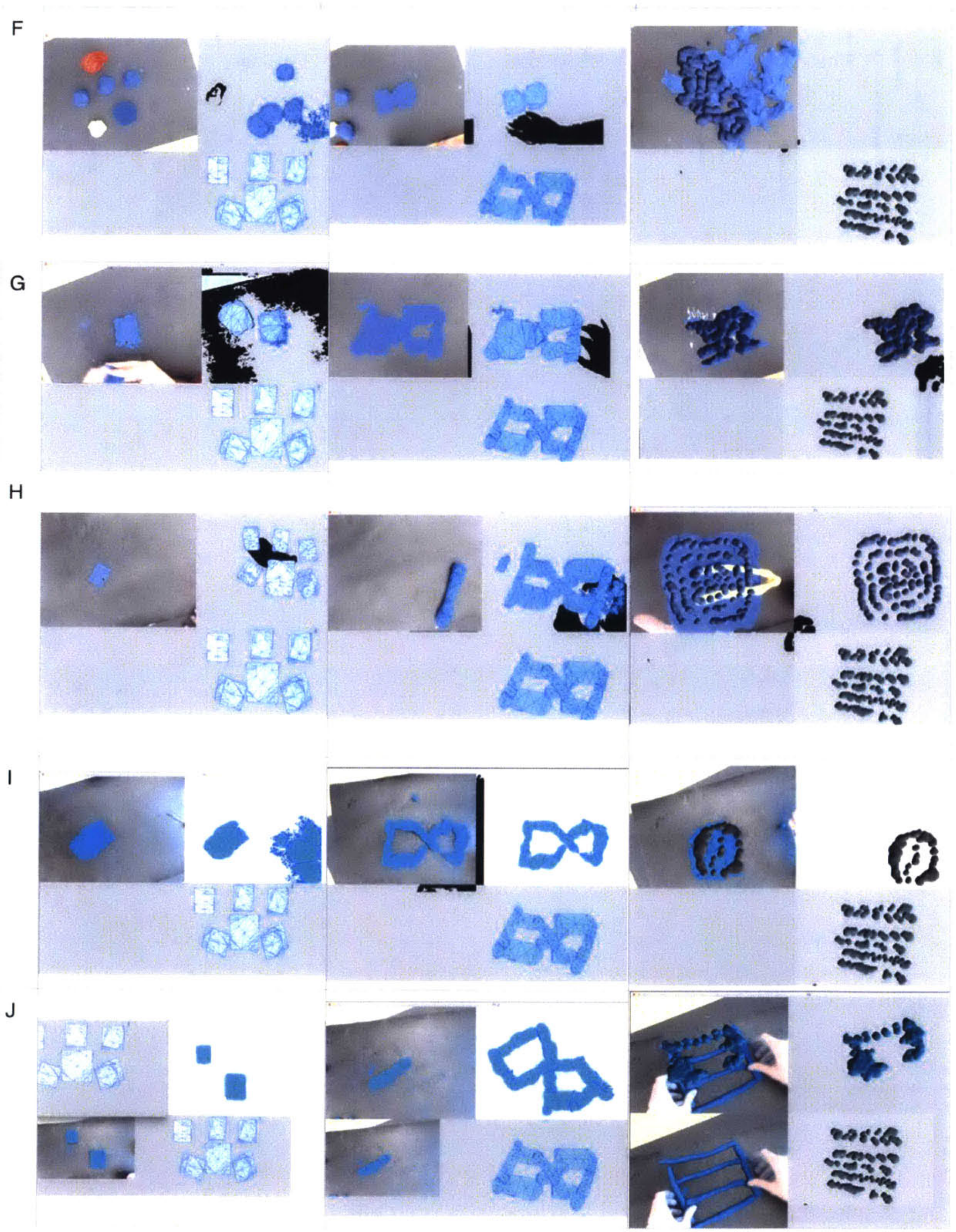
2.2 Data



Subject	Time P1	Time P1.5	Time P2	Prior Experience	Most difficult version?	Screen used more?
A. .	4.00	6.00	13.00	Uncomfortable w/ 3Dtransformations. Stong experience in Plan and Section from Architecture	3rd	Both
B.	6.00	skipped	4.00	Yes	3rd, didn't have a clear outline	White
C.	6.00	skipped	4.00	Architecture, software and games	3rd	White
D.	12.00 (misunderstanding)	skipped	5.00	Architectural Visualization	1st one. 3rd offered more "direct feedback"	White
E.	6.00	4.00	7.15	Yes, Gaming	3rd one. Planar easier	Both
F.	7.36	skipped	5.40	Yes.	1st one.	Both
G.	9.20		7.50	No.	P1.5. 1st was most frustrating.	
H.	3.12	6.06	5.08	Strong Experience with Photography Darkrooms	none.	White
I. .	7.00	5.20	11.20	Unity, Gaming, VR	3rd, "hidden logic"	visible hand
J.	5.40	10.20	16.30	Film-Making	3rd one	Hand

Screen captures of final results, each letter represents one subject:

#	P1 Final	P1.5 Final	P2 Final
A			
B		**skipped**	
C		**skipped**	
D		**skipped**	
E			


F
G
H
I
J

2.3 Theory Testing and Rule Finding

I set up the interface with three general rules to be learned by the subjects:

1. The visuals would always track to the largest contour of a specific color (blue).
2. Open holes within a shape would be triangulated into one large shape if the outlining contour is closed.
3. The top right screen can record images.

Subjects search for information when expectations are violated or a unique discovery is made by the subject. Similar findings suggests that self-made interventions contribute more to learning association than simply observing interventions.[34] The results also confirmed the importance of subject-driven interventions. For example, even though I would show users the recording property of the upper right screen few would internalize and use this property unless I explained it several times. This could be because it was something that they had observed, but didn't have the autonomy to try themselves, so it was often ignored. Instead, subjects relied mostly on tactics which they gained from their own self-interventions, especially those which were not anticipated and encountered through their own interventions.

Specific Theory-Testing:

Each person navigated through possible solutions to the task using different patterns of information-search.In some cases subjects would create forms in the clay which they believed to be a single solution to finding the task image. Subjects formed theories and tested these using clay (see Figure 18), ruling out a particular form of clay if it didn't work. This shows that the versatility of materials like clay was particularly conducive to testing out possible ideas, similar to sketching on pen and paper. Since it can easily change form it facilitates a process of testing, trying, changing and reiterating.

---

[34] Sobel, David M., and Tamar Kushnir. "The importance of decision making in causal learning from interventions." Memory & Cognition 34.2 (2006): 411-419.

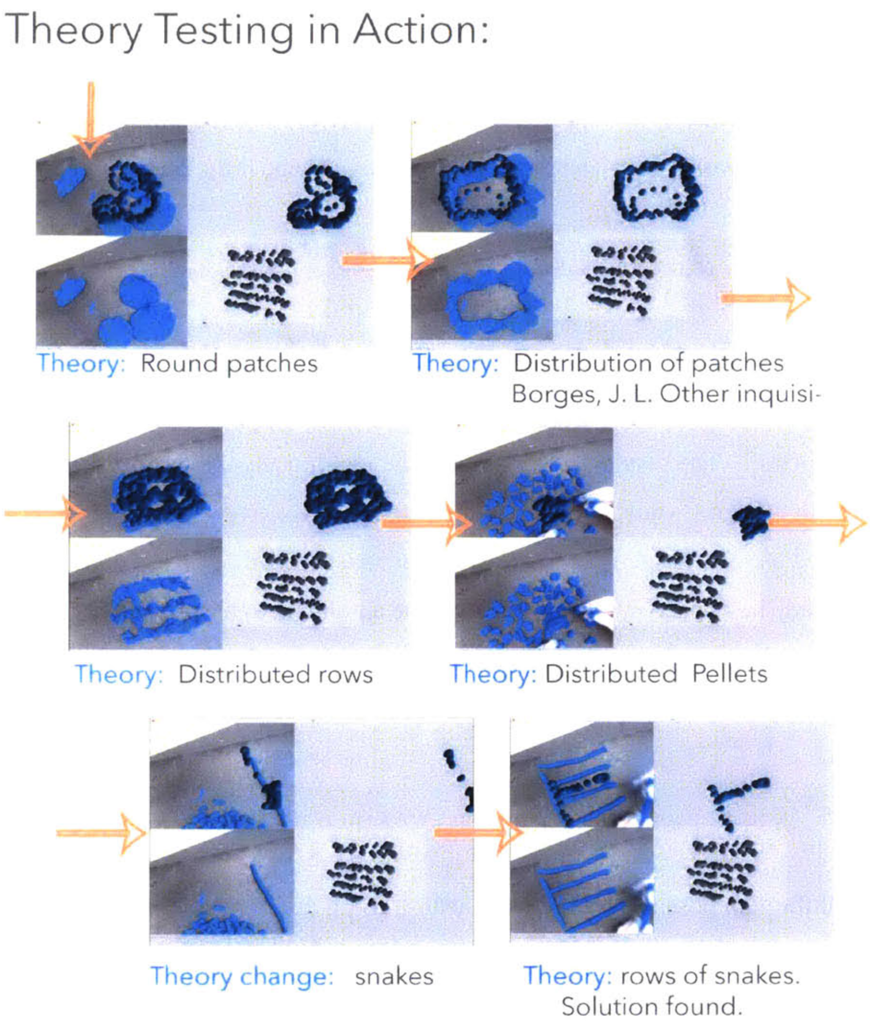


Figure 18: Navigating Towards Solutions using theories.

General Rule Finding:

Not all processing of interaction was as direct as "theory-testing" for specific forms. In some cases exploration was not as structured as trying one shape and ruling it out, rather there was a general rule assumption regardless of specific clay shapes. Even if that general assumption was incorrect technically, subjects could still employ that rule and achieve something close to the target. For example, 6 out of 10 subjects at some point tested the theory that a heightmap was at work (See Figure 19), and thought that the camera was a special sensor for height mapping (depth sensor). Of the 6 subjects who at some point assumed the height map theory, 2 of these maintained this theory even after interactions would have ruled this out.

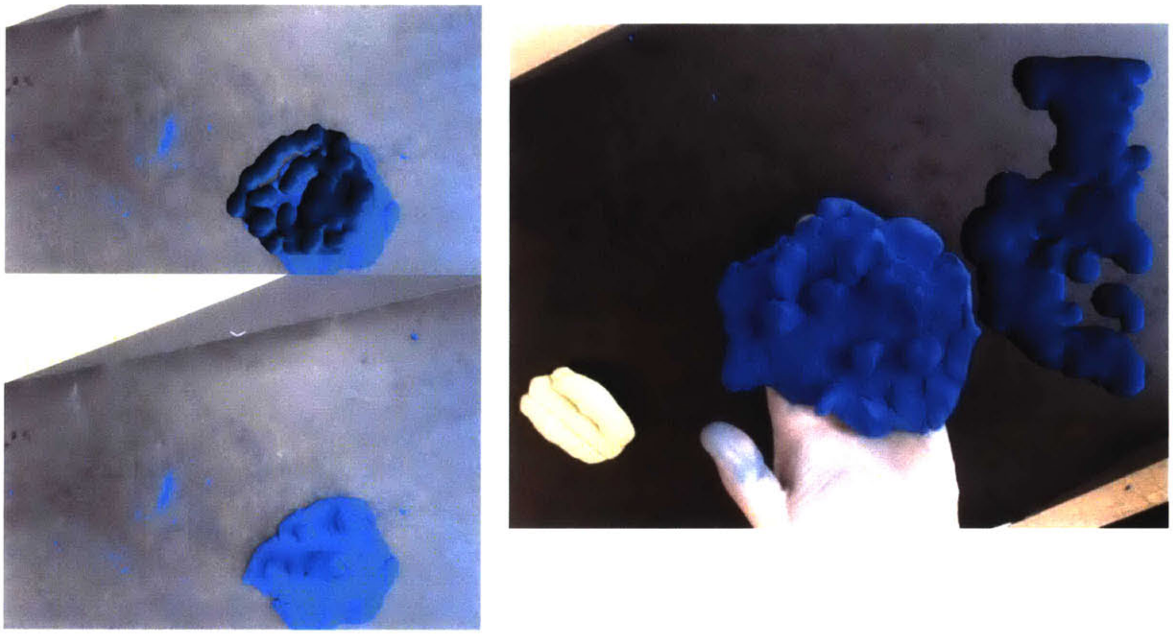


Figure 19: Assumption of a Height-Map.

2.4 Discovery: Transforming a discovery into a Tactic

> *"A true evolution of the form happens when a characteristic is turned into an attribute and given new values."*[35]
>
> -M. Yaghan, 2010

In several cases, when there was unexpected behavior in the interaction, subjects would use this found discovery as a tactic or tool for achieving the target image. For example, in one case, the subject discovered that a form with a hole in it (see Figure 20, left) would be represented as one filled polygon and subsequently chose to use this as a way achieve full squares as in the target image (see Figure 20, right).

---

[35] Yaghan, M. "The evolution of architectural forms through computer visualisation: muqarnas example." Electronic Visualisation and the Arts (EVA 2010) 3 (2010, page 113).

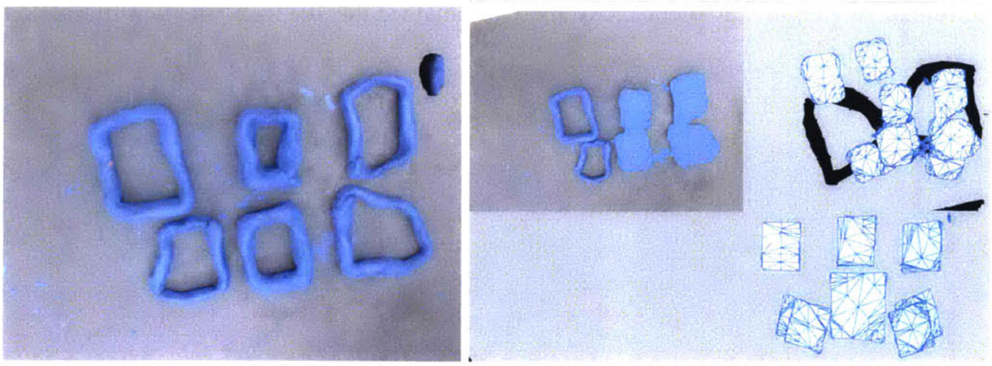


Figure 20: (Right) forms made purposefully with holes, (Left) screen capture of subject's view

In another example, the subject discovered that the voxel shape would rotate positions depending on where in the scene the contour was located (see Figure 21). They then chose to use this rotation mapping as a way to make dispersed lines similar to the target image, even though there were many other ways they could have found a solution.

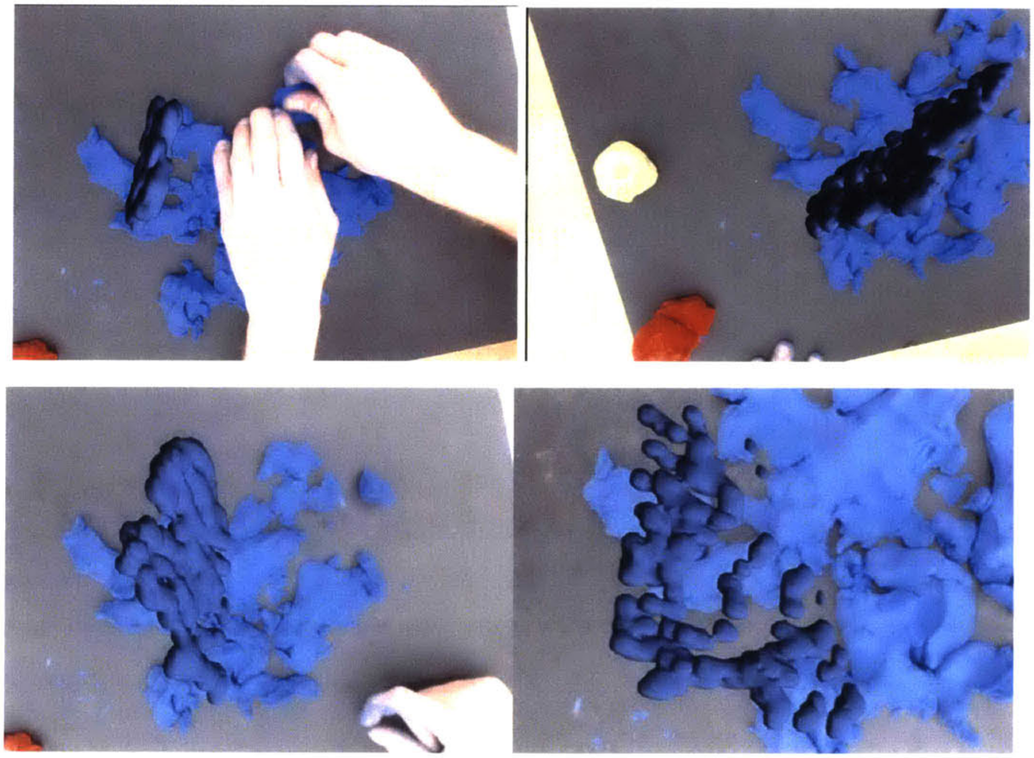


Figure 21: Use of discovered rotation-mapping in final solution.

In a third example, the subject was able to discover that, due to the fact that the size of the contour corresponds to sizes in the visual forms, they wouldn't have to actually physically make a square larger, they could instead, hold the smaller square closer up to the camera into order make it larger (see Figure 22) and record that moment. Arguably this could be due to the fact that this particular subject had experience with darkroom photography and was used to a light-capture device being held above a photograph. Factoring in the subject's experience, this is a mapping of form-manipulation specific to that subject's discoveries.

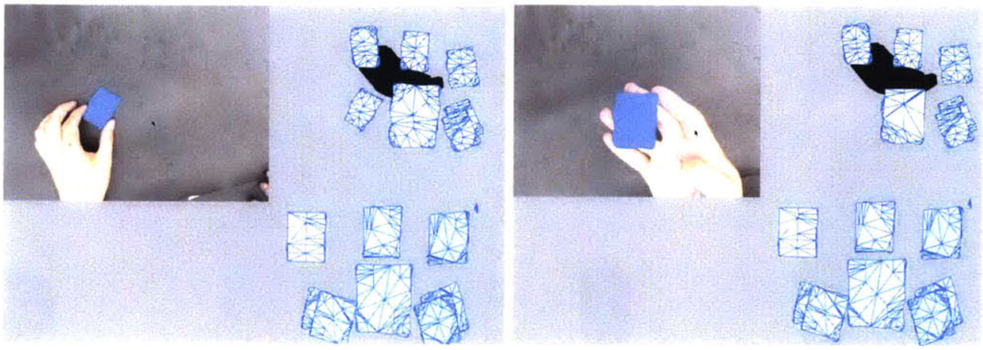


Figure 22: Subject makes the larger square by holding clay closer to the camera.

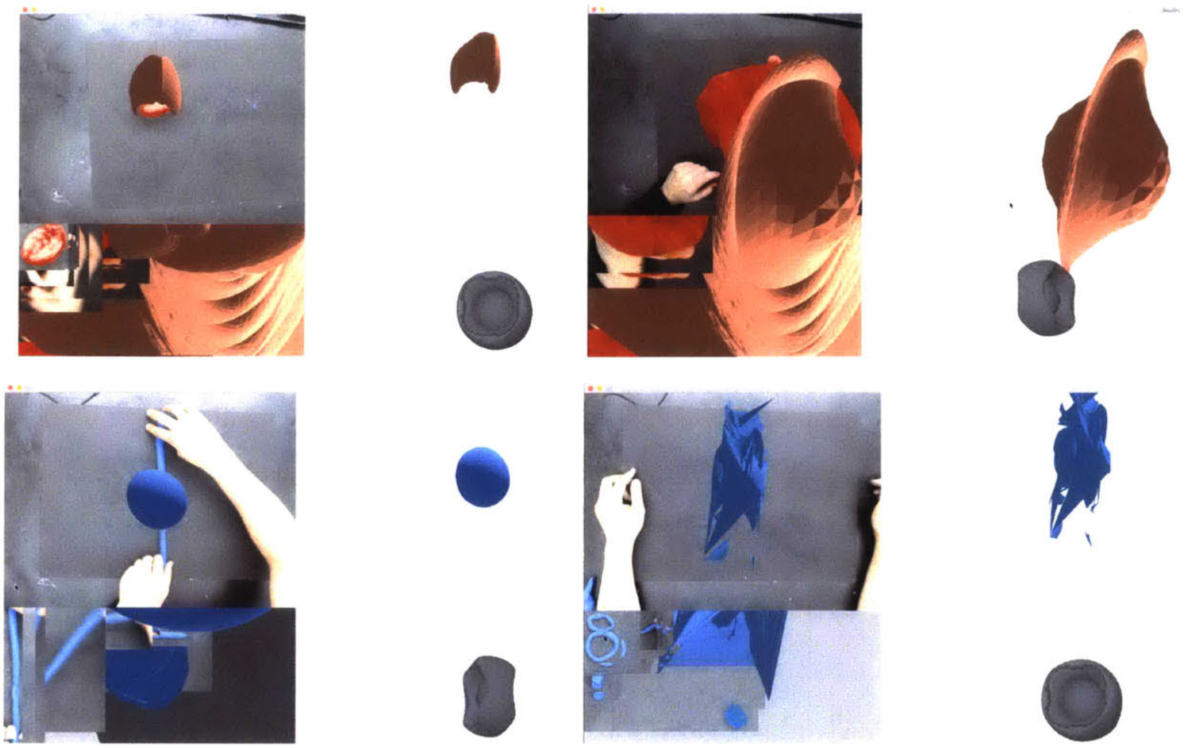


Figure 23: The erratic inflation behavior of version 3.

In the case of version 3, randomization was often assumed. Here the only rule was that the average size of the clay was correlated to size of the mesh, and making the clay very small would reset the physics dynamics. There is an amount of change which becomes more than the subject can find useful in a given set of interactions. This could be because there are dynamics which are extraneous to events, or because the behaviors aren't tightly correlated enough. While the rule was still found by all, there wasn't an ability to actually use that rule as a tactic because the resulting behaviors would continue without being correlated (see Figure 23). This needs to be further examined for more conclusive results.

Out of 5 separate subjects tested, 3 discovered the rule that occluding the contour or making it very small would reset the inflation physics. However overall, controllability was rarely gained, except in one subject. Mostly the subjects concluded that the behavior was too erratic or complex to control. One student of architecture found this was useful for creative form-finding. It was also interesting to observe the imagination of subjects in guessing about what the behavior represented: "morphing lookes like cell-fission" and "walnut shell implodes into something." The intention was to purposefully go against expected ideas of how the visuals would behave and see how far against expectation could still be controllable and yield a general "rule" about correlations in the subject's imagination. Studies about how infants perform exploratory information-search processes suggests that there is a point of too much information:

> *"Infants do not always prefer information that is more surprising or more complex: They appear to like things that they think they can learn. Thus, for instance, infants prefer to attend to patterns of linguistic input that are predictable enough to permit learning and to sequences of stimuli that are neither too predictable nor too unpredictable"*
> *- Shultz, 2015, Science p.43*

Further investigation into expectation-violation,[36] and why degrees of violation have an effect on interest and exploration, would need to be done before anything conclusive can be drawn about this in the case of adults and interfaces.

2.5 Limitations

☐Population Bias:

Graduate Students at MIT are a biased population for studying problem solving. All the subjects were in either in Comparative Media Studies, or in the Department of Architecture and Planning, (the one subject who was not from MIT was a computer programmer who had studied neuroscience so may be considered of the same population). The subject's advanced understanding of plan and section, gaming and film, and the relationship between a drawing and a 3D representation influence their ease of problem-solving across spatial transformations. More robust conclusions could be drawn from a more diverse population sample.

Experimenter Bias:

Naming and language, style of description all influenced subsequent assumptions about the program. Knowing intimately what the program would do may have led to unintended facial expressions, comments or body language while surveying the subjects.These were kept to a minimum, but likely unavoidable. Subjects also were asking questions during the experiment, which crucially revealed assumptions of how intentional the program was. For example, "I think this is a glitch," or "I don't think there is any way to get there [to the target form]." In these cases I noted their disbeliefs, but needed to persuade that there is a underlying "reason," otherwise the experiment would be ended. Subjects seemed

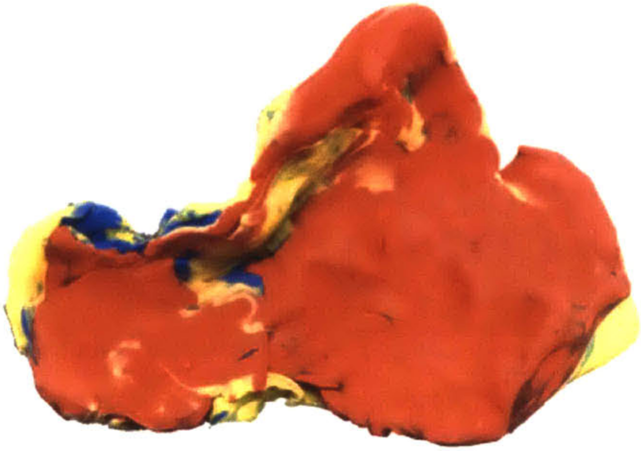
---

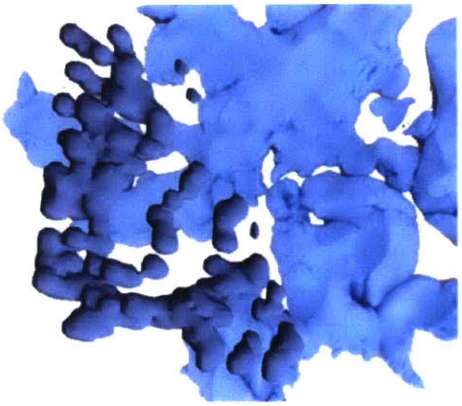
[36] Schulz, Laura. "Infants explore the unexpected." Science 348.6230 (2015): 42-43.

to assume there was a hidden logic, which was necessary to prevent subjects from either giving up, or playfully shaping clay at random and being indiscriminate about the effects. In subsequent studies these verbal exchanges, which significantly influence internal narrative, would be eliminated and only instructions in the beginning would be permitted. If there was an opportunity to test more subjects than experiments could have ended at the points where subjects decide to disbelieve that there is a relevant cause. Without an original assumption that there is *some* correlation subjects will not orient themselves towards the solution in nearly as much time. This was most evident in the case of Program 1 (P1), the non-visible case. For 3 of the subjects I needed to verbally reiterate that the screen was reacting to the clay in some way, after starting the experiment, otherwise (potentially due to language barriers or misunderstanding) they would assume that I was interested in the shape of the actual clay.

Technological Imperfections:

Lighting Conditions can impair the performance of the tracking, leading to actual glitches during interactions. The problem with this is mainly that it may have biased or influenced subjects towards believing the glitches were intentional, or that intended mappings were glitches and part of what they needed to understand. The augmented visuals tend to jitter due to the incoming video stream and the per-frame processing. In the future this could be refined using a smoothing algorithm, such as Kalman Filtering.





2.6 Conclusions

In observing how users make the target imagery, it was noticeable that they weren't relying or expecting the imagery to be realistically matched to the clay's form (visual fidelity), rather that correlations between changes in the clay and the imagery were more important to their sense of coherency, immersion and ability to control of the visual forms. My claim is that subjects discover their own situated mappings between physical and virtual interactions and that using a physical material, like clay, can facilitate the making and observation of these mappings. These mappings tended to occur after an intervention led to something unanticipated or surprising to the subject. These surprising events seem to be thought of as a discovery by the subjects, and consequently a useful tool (an affordance). In this way affordances emerge out of exploratory search and are structured on subject-driven interactions.

The use of computer-mediation is a plausible way to further understand exploratory information search and how that could lead to tool-making and tool-use at large. It also reveals the importance of the unexpected or imperfect, in its potential to be generative of creative use (affordances) within interfacing. The unexpected or erroneous moments that lead to a discovery are often followed by utilizing that property as a feature. In the process of exploratory search, unexpected qualities are not errors, rather they are incorporated into a logic of possible interactions.

Moments where an imperfection or unexpected behavior becomes a coherent discovery and subsequently a tactic or ability, hints at the adaptive dynamics by which humans have come to make "things" out of "stuff"[37] in the world to begin with. The goal then in designing interactive systems is to provide an equivalently "raw" yet coherent "stuff" out of which situated mappings can be built by the user. Here, the raw "stuff" is visual information on the computer screen, correlated to a material (CMMI).

The use of versatile materials and computer-mediated visual stimuli as a "raw" (less constrained) virtual-material suggests a different approach for structuring (destructuring) interfacing, in which an emphasis is placed on the subjects sequence of decision making through interaction and exploration, rather than on the formation of a learnable set of interactions in advance. Subject driven interactions could directly inform the development of affordances and constraints within a specific visual interface. Possible ways to achieve this are through using computer-mediation methods in real-time, or by incorporating machine learning (Adaptive Interfacing).

---

[37] Adelson, Edward H. "On seeing stuff: the perception of materials by humans and machines." *Photonics West 2001-electronic imaging*. International Society for Optics and Photonics, 2001. Page 1.

Chapter 3: Towards Adaptive Interfacing

3. 1 Future Steps

My experiments could be improved in the following ways. Tests could be run in less biased conditions and with more diverse subjects. Other correlation-based tests that use different sensory input (sound) have been considered. In addition to being informative, tests could be created in such a way that they could run off of browser-based platforms, such as Amazon Mechanical Turk (see Figure 24). Another element would be to test this condition in a multiple-user circumstance, to see how group dynamics, such as collaboration, might influence the subject-driven intervention process and the generation of affordances.

The use of devices such as an infrared input device, a head mounted display or head-tracker could be interesting and one could also comparatively evaluate the results to using a webcam only. Statistical Inference to detect and transfer material behaviors, or to register when a given pattern of interaction (gesture) has occurred multiple times, would allow one to push this type of computer-mediated material-interaction towards a more complex adaptive interfacing dynamic. In general, more complex underlying rules could be tested following the use of other computer vision algorithms such as feature extraction, as opposed to the somewhat simple blob detection.

Most importantly is the implementation of a better control test. A full control test would need to imitate the clay accurately, visually (texturally and in terms of physics behaviors) so as to analyze how users experience this accurate representation in comparison to the tests which deviate from exact replication. This is still challenging to do using only a single camera but is possible.

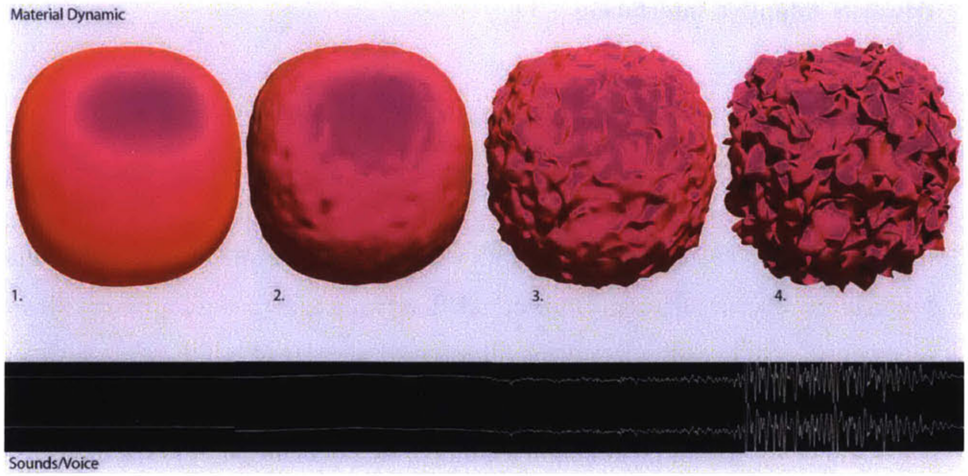


Figure 24: Modulations in the visual appearance of a material are based on the amplitude/inflections of the voice, in order to create a mapping between a virtual dynamic to a variable which is subject-controllable, but immaterial.

3.2 Towards Fully Adaptive Interfacing

The criterion for defining a "good" interface could be the utility that emerges from subject-driven and subject-specific interactions, which were possibly unintended by the designer. Subjects should be able to find channels of use and reuse that are both an unexpected discovery to them, as well as to the original intentions of the designer. This suggests that the interactions have a flexible underlying logic that allows for the space of possible interactions to develop over time. Interfaces could leverage the unexpected yet correlated events, motivating and encouraging subjects to engage in the exploration of its possibility space. An adaptive interface implies that both the subject and the interface interactions are adjusting, learning and adapting over time.

While how to do this remains an open question, there are some possible ways of implementing an approach to this. One method is to think of affordances as being unconstrained, or as a 'raw material' and allow definitions to happen based on the statistics of use and interaction. Allowing for imperfect interactions, or allotted time when a given software can switch modes from being constrained for a certain purpose, to being in a "raw" state (without intention or exploratory) are both possible applied methods of considering interaction as an adaptive medium.

One could use statistical learning algorithms to allow mappings to occur out of the frequency of interaction patterns with which the subjects naturally tend to engage. This would require longer trials of playing with the interface, but is another plausible method for implementing "adaptive interfacing," as the mappings would emerge from the statistics of interaction over time. The use of unsupervised learning algorithms, such as the self-organizing map (Kohonen Network) or probabilistic Bayes Nets, would be interesting both for the purpose of classifying and inferring patterns of interaction and for furthering emergent mapping. This means, allowing changes in the virtual forms to be based on a subject-defined set of interactions.

The videos of interaction generated in the process of performing this study could be utilized to train a network on the subject's patterns of interaction. In this case, classified pixels would be data input. Finally, it has already been proved that models can learn to predict physics behaviors from video,[38] so this would be very relevant and interesting to apply to live video streams, for behavioral (physics) feature transfer, augmented graphics and gaming environments.

Bibliography:

Adelson, Edward H. "On seeing stuff: the perception of materials by humans and machines." *Photonics West 2001-electronic imaging.* International Society for Optics and Photonics, 2001.

A.M.R. Agur and M.J.Lee. Grant's Atlas of Anatomy. Lippincott Williams and Wilking, 10th Edition, 1999

Biocca, Frank, and Ben Delaney. "Immersive virtual reality technology."*Communication in the age of virtual reality* (1995): 57-124.

Bratton, B. "The Convergence of Architecture and Interface Design." *interactions* (2008): 20-27.

Conn, C., Lanier, J., Minsky, M., Fisher, S., & Druin, A. (1989, July). Virtual environments and interactivity: Windows to the future. In *ACM SIGGRAPH Computer Graphics* (Vol. 23, No. 5, pp. 7-18). ACM.

Dehaene, Stanislas, et al. "How learning to read changes the cortical networks for vision and language." science 330.6009 (2010): 1359-1364.

Dobson, Velma, and Davida Y. Teller. "Visual acuity in human infants: a review and comparison of behavioral and electrophysiological studies." *Vision Research* 18.11 (1978): 1469-1483.

Ehrsson, H. Henrik, Nicholas P. Holmes, and Richard E. Passingham. "Touching a rubber hand: feeling of body ownership is associated with activity in multisensory brain areas." *The Journal of Neuroscience* 25.45 (2005): 10564-10573.

Eimer, Martin. "Multisensory integration: how visual experience shapes spatial perception." *Current biology* 14.3 (2004): R115-R117.

Evers, Vanessa, and Donald Day. "The role of culture in interface acceptance." *Human-Computer Interaction. INTERACT '97.* Springer US, 1997.

---

[38] Wu, Jiajun, et al. "Galileo: Perceiving Physical Object Properties by Integrating a Physics Engine with Deep Learning." Advances in Neural Information Processing Systems. 2015.

Graziano, Michael SA, Gregory S. Yap, and Charles G. Gross. "Coding of visual space by premotor neurons." *Science-New York, then Washington-* (1994): 1054-1054. Cognitive Science 31.3 (2007): 441-480.

Griffiths, Thomas L., and Joshua B. Tenenbaum. "From mere coincidences to meaningful discoveries." *Cognition* 103.2 (2007): 180-226.

Griffiths, Thomas L., and Michael L. Kalish. "Language evolution by iterated learning with Bayesian agents."

Guan, Haiying. *Vision-based Three-dimensional Hand Posture Estimation Using Hierarchical-ISOSOM.* ProQuest, 2007.

Harraway, Donna. "Situated knowledges: The science question in feminism and the privilege of partial perspective." *Feminist studies* 14.3 (1988): 575-599.

Held, Richard, and Alan Hein. "Movement-produced stimulation in the development of visually guided behavior." *Journal of comparative and physiological psychology* 56.5 (1963): 872.

Henrich, Joseph. *The secret of our success: How culture is driving human evolution, domesticating our species, and making us smarter*. Princeton University Press, 2015.

Kemmerer, David. *Cognitive neuroscience of language*. Psychology Press, 2014.
Lederman, Susan J., and Roberta L. Klatzky. "Extracting object properties through haptic exploration." *Acta psychologica* 84.1 (1993): 29-40.

Marr, D.; Poggio, T. (1976). "From Understanding Computation to Understanding Neural Circuitry". Artificial Intelligence Laboratory. A.I. Memo. Massachusetts Institute of Technology. AIM-357.

McLuhan, Marshall. *Understanding media: The extensions of man.* MIT press, 1994.

Pirolli, Peter. "An elementary social information foraging model." *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2009.

Piper, Ben, Carlo Ratti, and Hiroshi Ishii. "Illuminating clay: a 3-D tangible interface for landscape analysis." *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 2002.

Poincaré, Henri. *The value of science: essential writings of Henri Poincaré.* Modern library, 2012.

Poincaré, Henri. "Space and Geometry." (2011).

Riva, G., F. Davide, and W. A. IJsselsteijn. "Being There: The experience of presence in mediated environments." Being there: Concepts, effects and measurement of user presence in synthetic environments 5 (2003).

Rosch, Eleanor. "Principles of categorization." *Concepts: core readings*(1999): 189-206.

Schulz, Laura. "Infants explore the unexpected." Science 348.6230 (2015): 42-43.

Shams, Ladan, and Robyn Kim. "Crossmodal influences on visual perception." *Physics of life reviews* 7.3 (2010): 269-284.

Sharan, Lavanya, Ruth Rosenholtz, and Edward H. Adelson. "Accuracy and speed of material categorization in real-world images." Journal of vision 14.9 (2014): 12-12.

Sinha, Durganand, and Pushpa Shukla. "Deprivation and development of skill for pictorial depth perception." *Journal of Cross-Cultural Psychology* 5.4 (1974): 434-450.

Sobel, David M., and Tamar Kushnir. "The importance of decision making in causal learning from interventions." Memory & Cognition 34.2 (2006): 411-419.

Svenonius, Elaine. *The intellectual foundation of information organization.* MIT press, 2000.

Tabin, Clifford J. "Why we have (only) five fingers per hand: Hox genes and the evolution of paired limbs." Development 116.2 (1992): 289-296.

Warner, Julian. *Human information retrieval.* Cambridge, MA: MIT Press, 2010.

Winawer, Jonathan, et al. "Russian blues reveal effects of language on color discrimination." *Proceedings of the National Academy of Sciences* 104.19 (2007): 7780-7785.

Wu, Jiajun, et al. "Galileo: Perceiving Physical Object Properties by Integrating a Physics Engine with Deep Learning." Advances in Neural Information Processing Systems. 2015.

Yaghan, M. "The evolution of architectural forms through computer visualisation: muqarnas example." Electronic Visualisation and the Arts (EVA 2010) 3 (2010).