

# Towards User-Defined Multi-Touch Gestures for 3D Objects

**Sarah Buchanan**  
University of Central Florida  
4000 Central Florida Blvd.  
Orlando, FL 32816  
sarahb@cs.ucf.edu

**C.H. Bourke Floyd IV**  
JHT, Inc.  
2710 Discovery Dr.  
Orlando, FL 32826  
cfloyd@jht.com

**Will Holderness**  
University of Central Florida  
4000 Central Florida Blvd.  
Orlando, FL 32816  
wholdern@cs.ucf.edu

**Joseph J. LaViola Jr.**  
University of Central Florida  
4000 Central Florida Blvd.  
Orlando, FL 32816  
jjl@eecs.ucf.edu

## ABSTRACT

Although multi-touch interaction in 2D has become widespread on mobile devices, intuitive ways to interact with 3D objects has not been thoroughly explored. We present a study on natural and guided multi-touch interaction with 3D objects on a 2D multi-touch display. Specifically, we focus on interactions with 3D objects that have either rotational, tightening, or switching components on mechanisms that might be found in mechanical operation or training simulations. The results of our study led to the following contributions: a classification procedure for determining the category and nature of a gesture, an initial user-defined gesture set for multi-touch gestures applied to 3D objects, and user preferences with regards to metaphorical versus physical gestures.

## Author Keywords

Multi-touch; gestures; gesture recognition; referents.

## ACM Classification Keywords

H.5.2. Information Interfaces and Presentation: User Interfaces - *Interaction styles (e.g., commands, menus, forms, direct manipulation).*

## INTRODUCTION

With the widespread adoption of multi-touch devices, many users are familiar with standard multi-touch rotation, scaling, translation (RST) interactions for 2D direct manipulation of objects [14]. We envision multi-touch interaction to be incorporated into more 3D environments such as games, training, and simulations. Since there is no direct transition for standard 2D gestures to 3D direct manipulation, manipulating 3D objects on a multi-touch surface will require a new paradigm. In this paper, we begin to explore how familiarity with popular, metaphorical multi-touch gestures (e.g., a

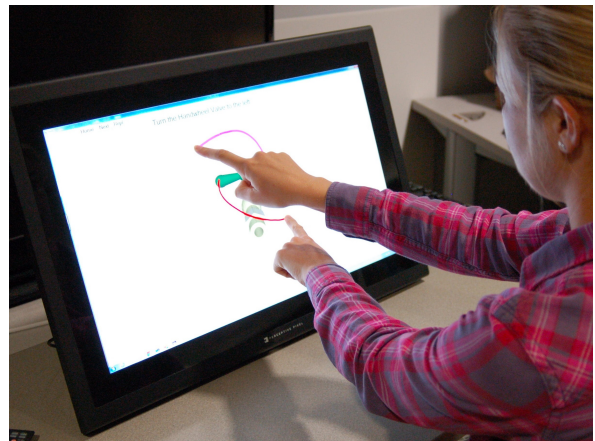


Figure 1: The Gesture Collection Apparatus and a user performing a two handed gesture.

swipe gesture to unlock) translates to interacting with 3D objects that afford physical actions (e.g., a knob affords turning) on a multi-touch device.

Wobbrock et al. defines a taxonomy of surface gestures containing four dimensions: form, nature, binding, and flow [14]. We will focus on distinguishing between user-defined metaphorical and physical gestures from the nature dimension. Metaphorical gestures are defined as gestures that act on, with, or like something else, while physical gestures are defined as gestures that should ostensibly have the same effect on a table with physical objects [14]. In this paper we expand on Wobbrock et al.'s classification by adding another category combining physical and metaphorical gestures called proxy gestures. We believe this research is a necessary precursor for future work on exploring how physical versus metaphorical multi-touch interactions can translate to learning real-world physical tasks. We specifically focus on interactions with 3D objects that have rotational, tightening, or switching components on mechanisms that might be found in mechanical equipment operations and training simulations.

We performed a user study to elicit user-defined gestures for manipulating 3D objects on multi-touch surfaces using a

study design established by Wobbrock et al [14]. The results indicate that some users naturally used previously learned metaphorical multi-touch gestures when interacting with the physical objects on multi-touch displays, while others instinctively used physical gestures. We were also able to show that with instruction to do so, users would switch from a metaphorical gesture to a physical one. Our user study led to the following research contributions:

- A classification procedure to categorize gestures and to determine the nature of a gesture (i.e., whether it is metaphorical or physical in nature, or a combination of both).
- A user defined gesture set for multi-touch gestures applied to 3D objects.
- User preferences with regards to metaphorical versus physical gestures.
- A comparison to Cohé et al.'s work that elicits user gestures for RST operations on a 3D cube [2].

## RELATED WORK

### Direct Manipulation and Physics Simulations

There has been much recent work on the manipulation of 3D objects on multi-touch surfaces. Direct manipulation is a widely explored strategy for this task, since the direct manipulation RST method has wide appeal in 2D contexts. Reisman et al. extends this common 2D paradigm to 3D by allowing direct manipulation of 3D objects with 3 or more touch points [10]. Hancock et al. also explores direct manipulation in [3] using one, two and three touch input techniques in shallow depth 3D. Physics based approaches have been applied by Wilson et al. [13] by creating solid proxy objects in the scene for each touch point. Physics based grasping behavior has also been explored by Wilson in a later work [12] where objects are manipulated by a stream of fluid particles. Cohé adapted the common mouse and keyboard transformation widgets to the tactile paradigm by creating a new 3D transformation widget tBox [1]. This body of work focuses on the direct manipulation of objects or widgets, whereas our work begins to explore how users intuitively act on, and how users prefer to act on, 3D objects from a certain domain. Our work will shed light on whether physics-based direct manipulation techniques are feasible and whether they would appeal to users in a training or operational environment.

### Metaphors for Manipulation

Gestures that act as metaphors for real world actions have been seen in previous work. Hancock et al. explored propagating behaviors to other objects in the scene by using metaphors, such as throwing a blanket object on top of another object to cover it in a texture [4]. Ruiz et al. elicit user-defined gestures for mobile interaction and find several themes similar to the ones we are exploring: actions that mimic normal use (such as putting a mobile phone to the ear for a motion gesture to answer a call), and real-world metaphors (such as placing a phone face down to hang up a phone call as you would have with a rotary phone) [11]. Kray et al. also discover user-defined gestures that act as metaphors for connecting mobile devices, displays and tabletops, such

as starting with two phones near each other and then pulling away to disconnect them [5]. Kurdyukova et al. investigate gestures for data transfer between iPads and other devices, finding that both experienced and inexperienced users rely on real-life metaphors when thinking of well-matching gestures [6].

### User-Defined Gestures

To create a user-defined, intuitive gesture set, Wobbrock et al. performed a study that elicited natural gestures from naive users [14]. Participants were presented with tasks to perform by showing the effect and asking the user to perform a gesture that would cause that effect. The users were asked to perform the gesture one-handed and then two-handed. It was determined that the number of fingers used for gestures was arbitrary for the same task and that users preferred one-handed to two-handed gestures. Because the study primarily focused on desktop operations and tasks, the final gesture set was heavily influenced by WIMP paradigms and yielded mainly metaphorical or symbolic gestures. They have since evaluated their user-defined gesture set against a gesture set created by designers and shown that the user-specified gesture set is easier for users to master [9].

There have since been many studies eliciting user-defined gestures based on Wobbrock's experimental design. Cohé conducted a user study to examine how users perform rotations, scaling, and translations on a 3D cube [2]. Our work is different than Cohé's in that we have added different objects and tasks to perform as well as two trials of the experiment. The first trial is similar to Cohé's in that it just asks the user to perform the gesture they think is appropriate for the given task. The second trial is different than Cohé's in that we ask the users to perform gestures as if they were manipulating the object in the real world. We believe that this is an important addition given our focus on multi-touch gestures for training and simulation applications in 3D environments. In addition to Wobbrock and Cohé's work, Ruiz et al. performed a study to elicit user-defined motion gestures for mobile interaction [11], Micire et al. studied user gestures for robot control and command in a 3D virtual environment [8], and Mauney et al. analyzed data from 9 different countries to determine cultural similarities and differences in user-defined gestures for touchscreen user interfaces [7].

### Classifications and Taxonomies

Wobbrock et al. also presented a taxonomy of surface gestures based on user behavior [14]. Based on a collection of gestures from participants, their taxonomy classifies gestures into four dimensions: form, nature, binding, and flow. [2] and [11] have adapted Wobbrock's original taxonomy to classify their specific gesture domains.

## USER STUDY

### Overview

The goal of our user-centered experimental design was to let users express their ideal gestures on objects that afford manipulation. The complex objects explored are valves, switches, tools, doors, and buttons that might be seen in mechanical

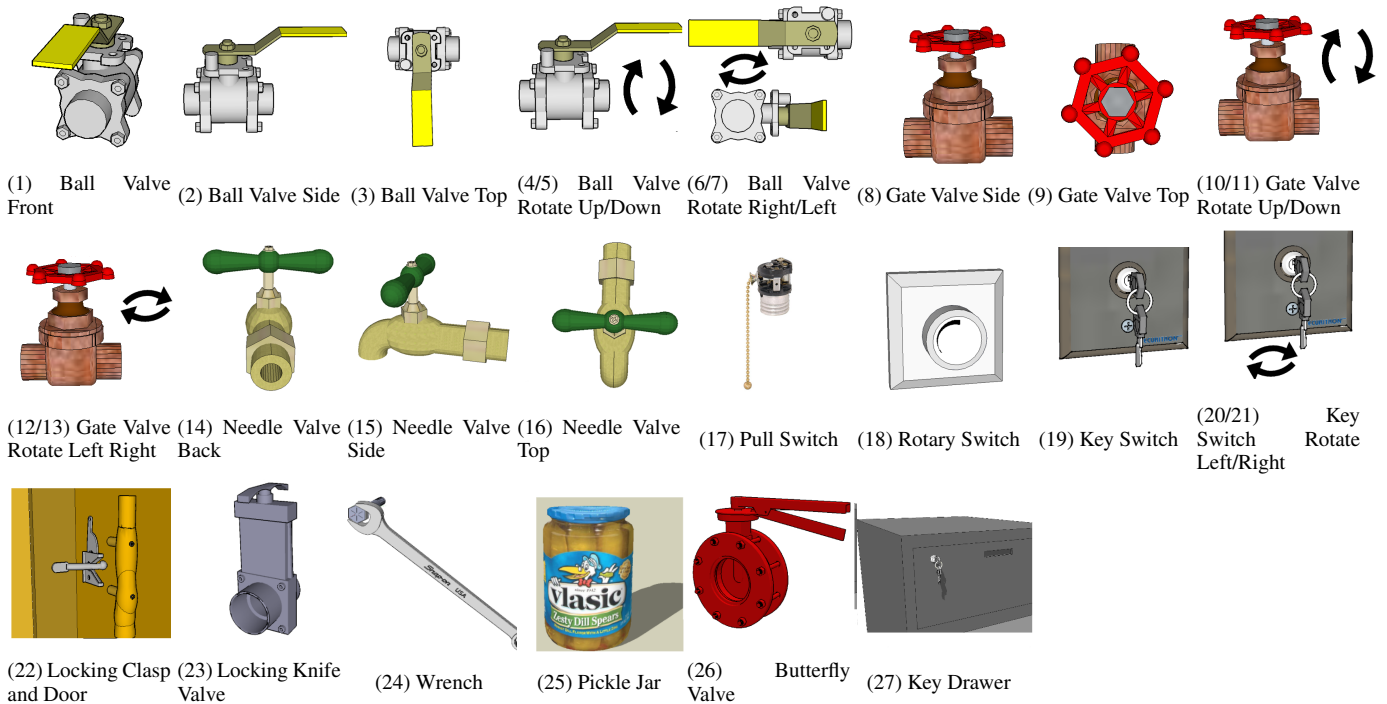


Figure 2: The 27 referents presented to participants in trials 1 and 2 of the user study.

Gesture	Abbrev.	# Fingers	# Hands
Dot	D	1	1
Straight Line	SL	1-3	1
Curved Line	CL	1-5	1
Semicircle	SC	1-3	1
Full Circle	FC	1-2	1
Spiral/ Multiple Circles	SP	1	1
Finger Turn	FT	2-5	2
Hold and Turn	HT	2-5	1
Pinch	PI	2	1
Pinch and Pull	PP	2-3	1
Hold and Drag	HD	2	2

Table 1: Observed gestures, as well as the observed range of fingers and hands used.

equipment operation and training. In order to remove bias from expected reactions to gestures, static images of these 3D objects were used with minimal feedback in the form of strokes drawn to the screen of their gesture path. All participants of the study stated they had previously owned multi-touch devices which caused inherent bias in what types of gestures participants naturally chose. Thus, we decided to go through two trials of the experiment. In the first trial, we asked the users to perform a task on each object using any gesture they felt was appropriate. In the second trial, we explicitly asked the users to perform gestures that they would use if this object was a physical object in the real world.

### Pilot Study and Gesture Primitives

A pilot study was conducted with 20 participants aged 19 to 26, with 13 males and 8 females using a 27" Perceptive Pixel LCD multi-touch display (PP display). During the pilot study,

we discovered a set of gesture primitives (shown in Table 1) that were applied to different objects in the current domain. Dot, straight line, curved line (shown in Figure 4B), semi-circle, full circle, and spiral are self-explanatory primitives when describing the motion of the contact points. If multiple fingers were used for those gestures, they all had to follow the same trajectory. Finger turn is a gesture where all contact points start in different locations and rotate around a central point (shown in Figures 1, 4A,D). Hold and turn is similar but contact points rotate around another contact point. Hold and drag is generally a two-handed gesture (however, one user performed a one-handed hold and drag) where one contact remains fixed and the other contacts move as a curved or straight line relative to it. The pinch and pull gesture is a pinch gesture followed by a straight line (shown in Figure 4C,F).

After the pilot was completed, to determine if these gesture primitives applied to more complex interactions that required possibly two hands or combinations of gestures as well as navigational tasks, we added new referents to the experiment that required compound operations (referents 22, 23, 25, 26, and 27 shown in Figure 2). We also added new navigational referents which required the user to change the perspective of the object by rotating the viewpoint (referents 4, 5, 6, 7, 10, 11, 12, 13, 20, and 21). The navigational referents are different than the other referents that require activating or turning on, in that they lend themselves to metaphorical gestures. Due to this, we will leave these referents out when comparing the nature results of the two experiment trials.

## Participants and Apparatus

There were 14 paid participants aged 18 to 29, all male. Two participants were left-handed and the remaining were right-handed. All participants owned a multi-touch device such as a phone, tablet or track-pad. Although all users were experienced with using multi-touch gestures on their devices, none had implemented a multi-touch application before. The experiment was conducted using a 27" Perceptive Pixel LCD multi-touch display (PP display). We developed an application (in C#) that displayed a static image of each referent next to an animated image of the task to perform (i.e., a valve opening) with a written description of the task (i.e. open this valve) as well. The application saved the users' gestures to a database. The data saved for each contact point included the timestamp, size, pressure, touch id, and coordinates which allowed for animated playback. In addition, each user was recorded on video.

## Procedure

Participants were asked to go through two trials of the experiment. For each trial the user was shown 27 referents (see Figure 2) accompanied by a task description on the screen (for example "Turn the gate valve to the left."). The scale of the objects on the screen was as close to real world scale as possible (shown in Figure 1). For the first trial, the participants were told to use whatever gestures they thought would be appropriate for accomplishing each task. For the second trial, the participants were told to use gestures as if each referent was a physical object in the real world. Immediately after each gesture was completed, participants rated their gesture for goodness and ease of use on a 7-point Likert scale. The total duration of the experiment was 30-45 minutes, which included both trials 1 and 2 and a post interview. After the 14 participants had completed the experiment with 27 referents and 2 trials, a total of 756 gestures were collected and classified.

## Classification Method

A systematic classification process was necessary to examine whether users interacted with the objects metaphorically, physically, or a combination of both. We define physical gestures in the same way as Wobbrock does; gestures that should ostensibly have the same effect on a table containing those physical objects. However, we distinguish ourselves somewhat from their definition by having the requirements that the gesture must use two or more fingers and the majority of the touches must make contact with the object. We make this distinction since if any of our objects was manipulated in the real world, those requirements would be necessary. Metaphorical gestures are then any gesture that uses 1 finger if they are also acting like the physical motion, or any other gesture that is not representative of the physical motion but is a metaphor in another way (such as a line in the direction the object should move). Proxy gestures are gestures that act like the physical motion but do not make contact with the object. The reasons we make these distinctions is first to prepare for implementing a physics simulation driven by projected contact with the objects, and second to mimic real world interactions for the purposes of training.

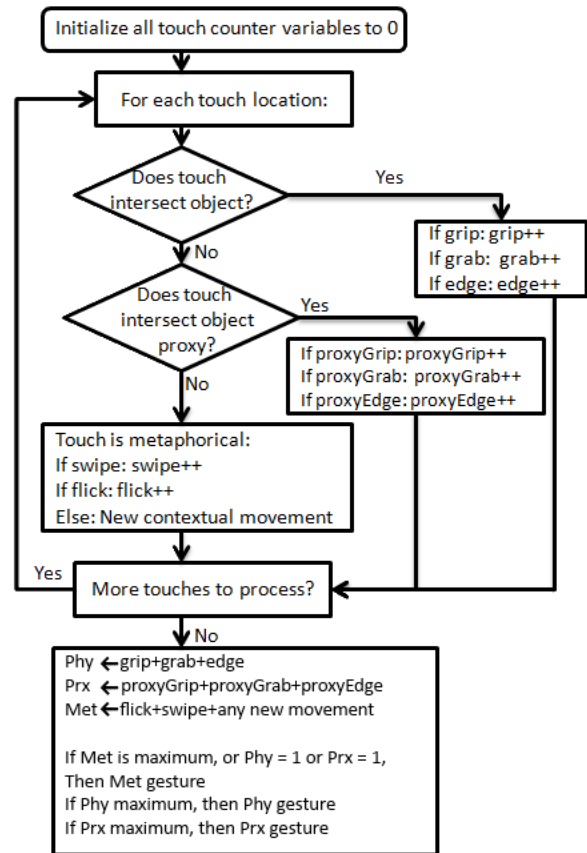


Figure 3: The nature classification process. The physical and proxy sub-categories are defined in Figure 4. The abbreviations Phy, Prx, and Met were used for Physical, Proxy, and Metaphorical gestures respectively.

The classification process (Figure 3) begins by examining the initial location and path of each touch and determining if it intersected the object. If the touch intersected the object and followed the path that was needed to apply force to the object, then it was classified as one of the physical gestures, either grip, grab, or edge (shown in Figure 4). A touch is classified as grip if its initial location intersects the object, edge if its initial location begins in empty space and then intersects the object, and acted like a grip, edge or grab motion, it was labeled a proxy grip, proxy grab, or proxy edge gesture. These proxy gestures mimic the physical motion done in real life, but since they do not make direct contact with the object they are metaphorical in nature. Thus we define proxy gestures as physically based metaphorical gestures.

Gestures that did not fall into the physical (Phy) or Proxy (Prx) categories, fell into the metaphorical (Met) category. We originally established a threshold for number of fingers to classify metaphorical gestures. Since real world motions could not be completed in our experiment with only 1 finger, the finger threshold was 1. The sub-categories for Phy, Prx and Met are shown in Table 2. After each touch location is

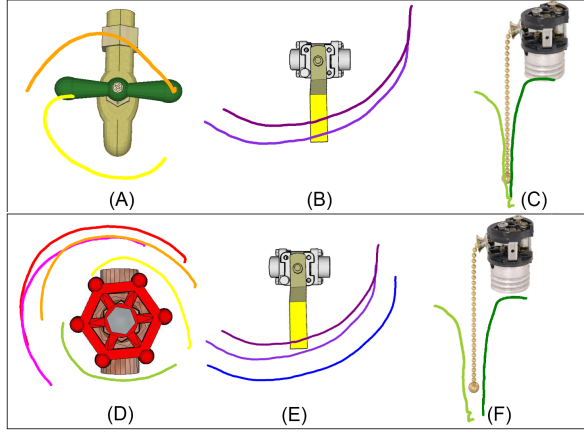


Figure 4: (A)-(C) demonstrate the Physical sub-categories from left to right: grip, edge, and grab. The gestures also correspond to 2 point finger turn, 2 finger curved line, and 2 finger pinch and pull. (D)-(F) demonstrate the Proxy sub-categories from left to right: proxy grip, proxy edge, and proxy grab. The gestures used are 5 point finger turn, 3 finger curved line, and 2 finger pinch and pull.

Metaphorical	Physical	Proxy	
Swipe(S)	Threshold Grip(TP)	Grip(P)	Proxy Grip(XP)
Hold/Drag(HD)	Threshold Grab(TB)	Grab(B)	Proxy Grab(XB)
Turn(N)	Threshold Edge(TE)	Edge(E)	Proxy Edge(XE)

Table 2: Sub classifications of gestures, as described in the Classification Method section.

classified, the sum of all of the Phy, Prx, and Met touches are compared to determine if the gesture is overall a Phy, Prx, or Met gesture. It is important to note that the abstract and symbolic gesture categories (presented in Wobbrock et al.’s taxonomy [14]) were left out of this classification method since no gestures in those categories were observed during our experiments.

## Experimental Results

### User-Defined Gesture Set

After all of the participants had run through the gesture collection experiment, each user’s gestures were classified according to the overall observed gestures shown in Table 1, the number of fingers and hands used, and the category (see Figure 4). Confirming Wobbrock et al.’s findings, our observations indicate that the number of fingers used is arbitrary in the interpreted gesture. However, the number of fingers is still important in interpreting the nature of the gesture indicated in Table 1. For instance, the number of fingers is still important as it can be indicative of more or less force or control. The gesture used by the majority for both trials for each referent is shown in Table 3. Although the interpreted gesture did not change for some objects, the physicality of the gesture increased. For example, for referent 1, Ball Valve Front, the average number of fingers used increased from 1 (SD = 0) in Trial 1 to 2.14 (SD = 1.04) in Trial 2. For many objects the gesture changed from a metaphorical gesture to a physical gesture. For example, referent 5, Gate Valve Side, had a

straight line Gesture in Trial 1 with average number of fingers = 1.3 (SD=0.57), which is a metaphorical gesture that does not apply to the real world. In Trial 2 referent 5 had a finger turn gesture with average number of fingers = 2.45 (SD=1.28) which is a physical gesture by our definition.

The gesture set found in the pilot remained the same through both trials of the experiment even though we added more complex objects and navigational tasks. It is worth noting that users did not create with new gesture primitives to accomplish these tasks that required compound operations (specifically referents, 22, 23, 26, and 27). Instead, we found users combined gesture primitives already defined to accomplish these tasks.

The user selected gestures for each referent were used to calculate an agreement score,  $A$ , that specifies the degree of participant agreement. This method was replicated from Wobbrock et al.’s prior work [14] and is defined as

$$A = \frac{\sum_{r \in R} \sum_{P_i \subseteq P_r} \left( \left| \frac{P_i}{P_r} \right| \right)^2}{|R|} \quad (1)$$

where  $r$  is a referent in the set of all referents  $R$ ,  $P_r$  is the set of proposed gestures for referent  $r$ , and  $P_i$  is a subset of identical gestures from  $P_r$ . The agreement scores indicate that there was more user consensus in gestures where the referent’s plane of motion was the same as the multi-touch surface (e.g. referent 9, Gate Valve Top, in Table 3), making the motion essentially 2D. For example, in Trial 2 referent 8, Gate Valve Front, has an obscured view where  $A = 0.34$ , whereas for referent 9, Gate Valve Top,  $A = 0.42$ . In addition, objects that were overall difficult to manipulate due to perspective or size had a low Agreement score regardless of the angle (e.g., in Trial 2 referents 20/21, Key Switch Rotate,  $A = 0.23$ ).

### Nature Dimensions

The results in Table 3 show, as expected, that the percentage of metaphorical (Met) gestures were higher during first pass of the experiment, and the percentage of proxy (Prx) or physical (Phy) gestures were higher during the second pass of the experiment. In Trial 1 there were 41.5% observed metaphorical gestures, 12.2% proxy gestures and 46.3% physical gestures, and in Trial 2 there were 10.2% metaphorical gestures, 24.8% proxy gestures and 65.0% physical gestures which shows an increase in physical gestures. However, the results do not show a majority of metaphorical gestures for the first pass of the experiment as we expected, indicating that some of the referents (such as those requiring compound operations) lent themselves to more physical gestures without the users being explicitly told to do so. If there was not a realistic way of physically interacting with the object on the screen due to viewing angle or awkwardness, users were forced to choose a proxy or metaphorical gesture, making the percentage of physical gestures lower than expected in Trial 2.

The gestures were also categorized according to the nature sub-categories shown in Table 2. Figure 5 shows the distribution of the different gesture sub-categories used per referent, which describe how users specifically made contact with the

	Referent	Gesture	Trial 1				Trial 2					
			Class	A	Met%	Prx%	Phy%	Class	A	Met%	Prx%	Phy%
1	Ball Valve Front	CL	TP	0.63	64.3	14.3	21.4	P	1.06	14.3	7.1	78.6
2	Ball Valve Side	CL	TP	0.53	50.0	21.4	28.6	E	0.80	14.3	21.4	64.3
3	Ball Valve Top	CL	TP	0.64	57.1	0	42.9	P	0.89	14.3	28.6	57.1
4/5	Ball Valve Rotate Up	SL	S	0.45	92.9	0	7.1	S	0.41	92.9	0	7.1
6/7	Ball Valve Rotate Right	SL	S	0.49	100	0	0	N	0.29	92.9	0	7.1
8	Gate Valve Side	FT	TP	0.34	28.6	21.4	50.0	XP	0.54	7.1	42.9	50.0
9	Gate Valve Top	FT	E	0.42	35.7	7.1	57.1	P	0.65	14.3	28.6	57.1
10/11	Gate Valve Rotate Up	SL	S	0.30	92.9	7.1	0	S	0.26	92.9	0	7.1
12/13	Gate Valve Rotate Right	SL(FT)	N	0.36	100	0	0	N	0.24	92.9	0	7.1
14	Needle Valve Back	FT	P	0.45	35.7	7.1	57.1	P	0.75	7.1	14.3	78.6
15	Needle Valve Side	FT	P	0.50	28.6	7.1	64.3	P	0.74	0	35.7	64.3
16	Needle Valve Top	FT	P	0.55	21.4	7.1	71.4	E	0.64	7.1	7.1	85.7
17	Pull Switch	SL	TP	0.88	57.1	35.7	7.14	XP	0.98	21.4	57.1	21.4
18	Rotary Switch	FT	TP	0.44	42.9	14.3	42.9	P	0.77	7.1	35.7	57.1
19	Key Switch	FT	XP	0.34	28.6	50	21.4	XP	0.53	0	50	50
20/21	Key Switch Rotate Left	SL(FT)	S	0.27	92.9	0	7.1	S	0.23	92.9	0	7.1
22	Locking Clasp / Door	CL	TP	0.44	92.9	0	7.1	P	0.28	42.9	7.1	50
		SL			42.9	0	57.1			7.1	7.1	85.7
23	Locking Knife Valve	FT	P	0.63	21.44	0,0	78.6	P	0.63	0	28.6	71.4
		SL			28.6	0	71.4			0	28.6	71.4
24	Wrench	CL	TP	0.63	50	7.1	42.9	E	0.66	0	0	100
25	Pickle Jar	FT	XP	0.33	42.9	28.6	21.4	XP	0.39	14.3	42.9	35.7
26	Butterfly Valve	PI	B	0.36	21.4	0,0	78.6	B	0.63	14.3	7.1	78.6
		CL			35.7	0	64.3			7.1	0	92.9
27	Key and Drawer	FT	TP	0.31	50	21.4	28.6	P	0.54	14.3	57.1	28.6
		SL			28.6	7.1	64.3			7.1	14.3	78.6

Table 3: The data collected for the 27 referents is shown. The Gesture column shows the majority gesture chosen in Trial 1 and Trial 2. Since the gesture values were mostly the same for both trials, only one is shown and if there was a difference in Trial 2 it is in parentheses. The gesture sub-classification is shown in the Class column. The Agreement scores (A), the percentage of Metaphorical (Met), Proxy (Prx), and Physical (Phy) gestures used, are also shown respectively.

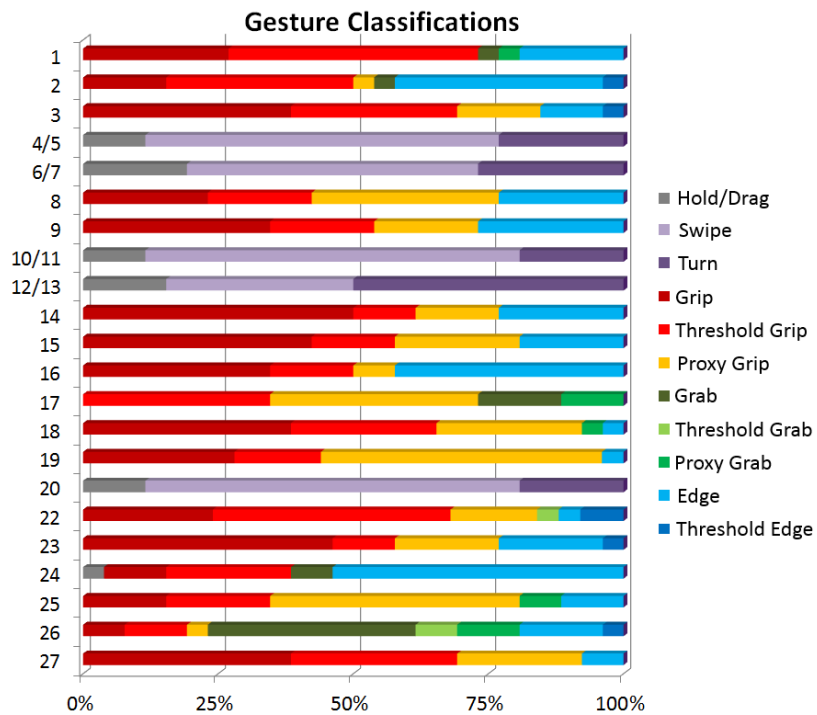


Figure 5: The distribution of the different gesture sub-categories used per referent over both trials, which describes how users specifically made contact with the object (defined in Table 2).

object. The results demonstrate that gripping, whether physical grip, proxy grip, or threshold grip, is the majority method of contact for all referents over both trials. Grip means that users specifically made contact with the parts of each referent they wished to apply force to as if they could drag that area. The second most common contact type was edge, where users made contact with space and then moved into the object's space as if they are applying force to the edge of that object. Although there were gestures made up of proxy edge components, none of the users performed a gesture that had a majority of proxy edge touches, which is why no proxy edge gestures are shown in Figure 5.

#### *Number of Fingers and Hands*

As we assumed, most users gestured with 1 or 2 fingers during Trial 1 and more than 2 fingers in the Trial 2. The average number of fingers across all referents was 1.89 in Trial 1, and 3.49 in Trial 2, which is significantly greater in Trial 2 ( $t_{12} = -5.99, p < 0.01$ ). The average number of hands in Trial 1 was 1.04 and in 1.07 in Trial 2.

#### *Gesture Rating*

Similar to Wobbrock et al.'s previous work, after the participants completed each gesture they rated the goodness and ease of use on a 7-point Likert scale. The first Likert item stated "The gesture I picked is a good match for [task here]" and the second stated "The gesture I picked is easy to perform." The Mann-Whitney tests showed that there were no significant differences in ratings between trials of the experiment. Again the ease of use ratings indicate that users did not perceive the physical gestures used in the second pass of the experiment to be more difficult to use than the metaphorical gestures used in the first pass of the experiment. This is interesting because symbolic or metaphorical gestures are considered simplifications and abstractions of real world actions and, therefore, considered easier to perform.

There were, however, significant differences between referents. For example, referent 8 (Gate Valve Side) had an average goodness rating of 5.93 (SD = 0.10) and referent 9 (Gate Valve Top) had an average goodness rating of 6.34 (SD = 1.07), which are significantly different ( $Z = -3.27, p < 0.05$ ). In addition, the ease of use ratings are significantly different ( $Z = -2.14, p < 0.05$ ), where referent 8 again has a lower mean ( $\bar{x} = 5.51, SD = 1.43$ ) than referent 9 ( $\bar{x} = 5.99, SD = 1.14$ ). The camera angle (in the previous case) as well as the difficulty of manipulating the referent (referent 19 versus 18 for example) play a part in how the users perceived their gestures' goodness and ease.

#### *Interviews*

After the participants finished the 2 trials they answered 3 interview questions before the experiment was complete. The interview questions were: (1) "Did you notice that your gestures changed in the second pass of the experiment when I said to treat the objects as real, physical objects?", (2) "What specifically changed about your gestures in the second pass?", (3) "Which gestures did you prefer - the ones used in the first pass or those in the second pass, and why?" The majority of users answered yes (11/14) to the first question. Those users specified either more fingers, pressure, or parts of hand as

the way their gestures changed in the second trial. One user stated that "At first I interacted as though it was a phone app, and then I incorporated more fingers." The other 3 participants were already using physical gestures and using more fingers and hands in the first pass of the experiment.

For the final question, the participants were evenly divided in their preferences, 6/14 (42.9%) indicated they preferred the first pass of the experiment mostly because they were easier to perform, and the physical gestures could be awkward. The other 6/14 participants indicated they preferred the second pass of the experiment. According to one user "the gestures were similar to habits I use every day", and another said "it's easier to do things as if you would in real life". 2/12 (14.3%) said that their preference depends upon the referent and the viewing angle. Although it may seem that physical gestures would correspond to physical objects, we assumed the majority would prefer the metaphorical gestures from the first pass, since there is inherent bias from the use of multi-touch phones and tablets. Thus, we were surprised that user preferences were evenly divided. One user stood out from the others stating that he used the Mac track-pad and because of this he "prefers multiple fingers because it's a more intuitive experience to turn a knob the way you would in real life." In addition, some users stated that it would depend on the object and the application. In particular, one user stated "It would be really cool to have a video game where you had to navigate the world by interacting with the picture like they were real objects. For simple interfaces, though, I would rather have simpler controls."

## **DISCUSSION**

### **Results**

We classified each gesture into the categories within the Nature dimension (Phy, Prx, or Met) as well as the sub-categories shown in Table 2. All gestures were observed as either in the metaphorical category or in one of the physical or proxy sub-categories. The number of fingers played an important role in distinguishing between metaphorical and physical gestures. For instance, a curved line with one finger to open a ball valve (Figure 2-1) was considered a metaphorical gesture since in the real world more force would usually be required, whereas a curved line with 2-5 fingers would be considered a physical gesture because it mimics the real world motion. This is not to say that more fingers always leads to a physical gesture. For instance, opening a gate valve (Figure 2-8) with a 3 finger straight line gesture would be considered a metaphorical gesture because that would not translate to real world movement whereas a multi-point finger turn gesture would.

In Trial 1 there were 41.5% observed metaphorical gestures, 12.2% proxy gestures and 46.3% physical gestures, and in Trial 2 there were 10.2% metaphorical gestures, 24.8% proxy gestures and 65.0% physical gestures which shows an increase in physical gestures. However, the results do not show a majority of metaphorical gestures for the first pass of the experiment as we expected. This indicates that some of the referents (such as those that required compound operations) naturally lent themselves to more physical gestures without

the users being explicitly told to do so. Viewing angle and the awkwardness of the object cause the percentage of physical to be lower in Trial 2 as well, since if there is not a realistic way of physically interacting with the screen, the user chose a proxy or metaphorical gesture.

### Related Work Comparison

Our results contain both similarities and differences with Cohé's study [2]. Both studies determine how users intuitively manipulate 3D objects, where Cohé uses a 3D cube alone and we use more complex objects. Although the objects are different in our study the results are similar and complementary to Cohé's results upon further investigation. Cohé examines three parameters to categorize gestures: form (number of fingers), initial point location (IPL), and finger trajectory. With this information Cohé determines the location of the IPL on the cube (e.g., corner, edge, face, or external to the cube) and defines a relationship between the IPL and the transformation. For instance, for rotating a cube the most common choice was an IPL on a face parallel to the transformation axis (TA) with a trajectory along the transformation direction (TD). From these fine grained observations, Cohé found several overall characteristics of gestures for rotating an object:

- Curved - the trajectory is curved.
- Straight - the trajectory is straight.
- Grab - the user picks a point on the cube surface and then moves the object.
- Push - the user begins their gesture and the cube moves after it has been pushed, or when the finger trajectory intersects an edge orthogonal to the TA.

From these general characteristics emerged four gesture categories that encompass all rotation gestures: CurvedAndPush, StraightAndPush, StraightAndGrab, and CurvedandGrab. Similarly, Cohé came up with the following characteristics for translation gestures: Push, Without object referent, Grab-Lateral, Grab-Pull and Grab-Push. Straight and curved were omitted since there were no curved gestures observed for translation.

Our classification method also examines form, IPL, and finger trajectory to determine what the gesture shape is (listed in Table 1), then categorizes each touch according to the process in Figure 3, and finally comes up with an overall determination if the gesture was Met, Prx, or Phy. Interestingly, we came up with similar categories for gesture classification. Our grip, edge, and proxy categories correspond to Cohé's grab, push, and without object referent categories respectively. We also added the category for a grab that represents an enclosing or pinching motion that someone would do to pinch or grab an object in the real world.

We believe our data from Trial 1 is similar to Cohé's for referents 4-7, 10-13, 20, and 21 since those referents are rotating the entire object's viewing angle, which is similar to rotating the bounding box of an object. However, in Trial 2 users were told to do gestures as they would in the real world, so the number of fingers used increased which is different than

Cohé's gesture categories that use primarily 1-2 fingers. Also in Trial 2, to perform a rotation, users would mimic picking up an object and rotating it which doesn't fit into Cohé's categories for rotations. For the remaining referents for Trial 1, the data was also similar to Cohé's since turning a valve or switch on is rotating or translating a particular part of that component. Again in Trial 2, these tasks would use more fingers and would not fit into Cohé's categories. It is interesting to note that Cohé defined all of the user's gestures as physical gestures. Whereas by our definition, all of the gestures performed by users in Cohé's study would be metaphorical since they would not perform the desired action in the real world.

### Design Implications

In Trial 1 of our study users intuitively used gestures similar to Cohé's to manipulate real world objects, thus verifying Cohé's work. However, when asked to treat the referents as real world objects users performed more physical gestures. In addition, when performing more physical gestures users consistently agreed that they were no more difficult to perform than their metaphorical counterparts. This is ideal for future uses of physical gestures in simulation or training environments. Users also were evenly divided on their preference for metaphorical gestures or physical gestures. It is also interesting that the viewing angle affected the perceived difficulty to perform a gesture and is something to keep in mind when designing these systems.

The results of our study shed light on the feasibility of designing a system that requires physical gestures as input. There would be two ways to develop such a system. The first approach would be to implement a gesture recognizer that would look for a certain gesture shape or shapes combined with initial touch locations. The referent context would be necessary to determine which gestures are acceptable for the specific context. The second approach would be to interface with a physics engine by projecting the touch locations into the scene to determine collisions with the objects. A grip touch point would project directly to the object, whereas an edge or grab touch point would need situational context. Supporting proxy gestures is recommended, especially for objects that are smaller than a user's hand or have awkward perspectives. Proxy gestures could be implemented by having invisible handles in the scene that the user's touch locations would make contact with. Adding number of finger or hand requirements, as well as pressure requirements, could also add realism to the scene. In a physics engine implementation, the number of fingers and amount of pressure could increase the amount of force applied to the object. However, some devices do not have pressure sensitivity, and have a limited number of touch inputs, so that would be a device dependent requirement.

### FUTURE WORK

In the future we would like to examine a richer set of 3D objects and whether training using physical gestures versus metaphorical gestures translates to better performance in a real world task. For instance, consider a mechanical or medical task that requires operations that take some time and precision to perform. Having a training environment that only re-



quires the user to click or do simple gestures to work through an exercise may not prepare them for the real world tasks as well as more realistic physical gestures. It would be interesting to see how pressure used or the time taken to perform such operations play a role in training. Perspective was also a challenge encountered in this study that would be interesting to explore. For example, if an object's plane of motion is the same as the multi-touch surface it is easier and more intuitive to manipulate than something that does not align with the viewing plane (such as referent 9, Gate Valve Top, versus referent 8, Gate Valve Side). Some users in our study chose to do the same gesture, regardless of perspective, and it would be interesting to see how users preferred to deal with alternate perspectives.

## CONCLUSION

We have presented a user study that explores how standard RST-style multi-touch interaction transitions to 3D objects that have rotational, tightening, and switching components. We have also described a procedure for classifying gestures as metaphorical, physical, or proxy. Our results show a large spread of metaphorical and physical gestures in Trial 1 when users were given no guidance. We believe this large spread is indicative of existing biases from previous multi-touch experiences leading to metaphorical gestures, combined with the domain of the referents lending themselves to physical gestures. However, once prompted to use gestures as if manipulating physical, real-life objects, the users increased the number of contact points and used more gestures that were physical in nature. The participants also found these physical gestures just as easy to perform as metaphorical gestures. Designers should take into consideration that users intuitively use the gestures they are most familiar with, so if they would like to elicit physical gestures there needs to be guidance in doing so. Other than intent, many other factors play a role in the gestures chosen, among them form factor and the perspective view of the referent. In addition to the gestures, there is also the intersection and interaction with the object to take into consideration when interpreting the gesture, as defined in the physical and proxy nature sub-categories presented.

## ACKNOWLEDGMENTS

This work is supported in part by JHT Inc. Award JHT12S0003, NSF CAREER award IIS-0845921, and NSF awards IIS-0856045 and CCF-1012056. We would also like to thank the members of ISUE lab for their support and the anonymous reviewers for their useful comments and feedback.

## REFERENCES

1. Cohé, A., Dècle, F., and Hachet, M. tbox: a 3d transformation widget designed for touch-screens. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, ACM (New York, NY, USA, 2011), 3005–3008.
2. Cohé, A., and Hachet, M. Understanding user gestures for manipulating 3d objects from touchscreen inputs. In *Proceedings of Graphics Interface 2012*, GI '12, Canadian Information Processing Society (Toronto, Ont., Canada, Canada, 2012), 157–164.
3. Hancock, M., Carpendale, S., and Cockburn, A. Shallow-depth 3d interaction: design and evaluation of one-, two- and three-touch techniques. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '07, ACM (New York, NY, USA, 2007), 1147–1156.
4. Hancock, M., ten Cate, T., and Carpendale, S. Sticky tools: full 6dof force-based interaction for multi-touch tables. In *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces*, ITS '09, ACM (New York, NY, USA, 2009), 133–140.
5. Kray, C., Nesbitt, D., Dawson, J., and Rohs, M. User-defined gestures for connecting mobile phones, public displays, and tabletops. In *Proceedings of the 12th international conference on Human computer interaction with mobile devices and services*, ACM (2010), 239–248.
6. Kurdyukova, E., Redlin, M., and André, E. Studying user-defined ipad gestures for interaction in multi-display environment. In *Proceedings of the 2012 ACM international conference on Intelligent User Interfaces*, IUI '12, ACM (New York, NY, USA, 2012), 93–96.
7. Mauney, D., Howarth, J., Wirtanen, A., and Capra, M. Cultural similarities and differences in user-defined gestures for touchscreen user interfaces. In *CHI '10 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '10, ACM (New York, NY, USA, 2010), 4015–4020.
8. Micire, M., Desai, M., Courtemanche, A., Tsui, K. M., and Yanco, H. A. Analysis of natural gestures for controlling robot teams on multi-touch tabletop surfaces. In *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces*, ITS '09, ACM (New York, NY, USA, 2009).
9. Morris, M. R., Wobbrock, J. O., and Wilson, A. D. Understanding users' preferences for surface gestures. In *Proceedings of graphics interface 2010*, Canadian Information Processing Society (2010), 261–268.
10. Reisman, J. L., Davidson, P. L., and Han, J. Y. A screen-space formulation for 2d and 3d direct manipulation. In *Proceedings of the 22nd annual ACM symposium on User interface software and technology*, UIST '09, ACM (New York, NY, USA, 2009), 69–78.
11. Ruiz, J., Li, Y., and Lank, E. User-defined motion gestures for mobile interaction. In *Proceedings of the 2011 annual conference on Human factors in computing systems*, ACM (2011), 197–206.
12. Wilson, A. D. Simulating grasping behavior on an imaging interactive surface. In *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces*, ITS '09, ACM (New York, NY, USA, 2009), 125–132.

13. Wilson, A. D., Izadi, S., Hilliges, O., Garcia-Mendoza, A., and Kirk, D. Bringing physics to the surface. In *Proceedings of the 21st annual ACM symposium on User interface software and technology*, UIST '08, ACM (New York, NY, USA, 2008), 67–76.
14. Wobbrock, J. O., Morris, M. R., and Wilson, A. D. User-defined gestures for surface computing. In *Proceedings of the 27th international conference on Human factors in computing systems*, CHI '09, ACM (New York, NY, USA, 2009).