Drag and Drop the Apple: The Semantic Weight of Words and Images in Touch-Based Interaction

Ilhan Aslan, Martin Murur, Verena Fuchsberger, Andrew Fugard, Manfred Tscheligi
Christian Doppler Laboratory for Contextual Interfaces, HCI & Usability Unit, ICT&S Center, University of Salzburg, Austria
firstname.lastname@sbg.ac.at

ABSTRACT
In this paper we report a user study to investigate the effect of semantic weight in a touch-based drag and drop task. The study was motivated by our own interest in exploring potential factors that influence touch behavior and supported by results in related neuroscience research. The question we intended to answer is: “Do people drag the representation of a smaller and lighter real world object (e.g. an apple) different than the representation of a heavier and larger real world object (e.g. a car)?” Participants were asked to perform a drag and drop task repeatedly on a tablet device. Dragged objects were the same physical size on screen, but represented real world objects that were either heavy and large or light and small. We studied two representation modalities (i.e. image and text). In both representation modalities, semantically heavier objects were dragged significantly faster than semantically lighter objects.

Author Keywords
Touch, Performance, Behavior, Vision, Semantics.

ACM Classification Keywords
H.5.2 Information Interfaces and Presentation: User Interfaces]

General Terms
Design, Experimentation, Human Factors, Measurement.

INTRODUCTION
For many years, HCI researchers have been examining the benefits of different touch-based interaction techniques. The analysis of these interactions are usually based on measures explaining a speed/accuracy trade off. So far, HCI research on performance analysis is only focussing on traditional spatial measures (e.g. position, distance and target size) to explain performance results. Designers benefit from this knowledge to improve the usability of their designs. For example, it is already known that interactive objects have to be larger in touch based interaction than they are for mouse-enabled interaction. Recent research has also studied how targets are acquired in touch-based interfaces [9, 10] and that on Smartphones target acquisition is systematically skewed depending on the position of targets.

In many cases, it is assumed that our finger and hand movement is guided solely by our conscious visual perception and therefore only spatial measures in interfaces are utilized for the analysis of user performance and behavior. It is common that “unimportant” details of a real interface that exhibit variations (i.e. textual and graphical representations that are associated with different meanings) are ignored; this may or may not be relevant for users’ performance. Researchers in neuroscience demonstrated the influence of visual presentations on the activation of motor tendencies related to semantic qualities of the objects that are presented. For example, reading the word “apple” will activate different motor tendencies than reading the word “grape”, because an apple is larger and heavier than a grape. Predicted effects exist for different categories of semantic qualities and perceptions (e.g. perceptions of weight, fragility, texture, surface friction and temperature [7]).

In this paper, we report a study to investigate the influence of those semantic qualities that users perceive during interaction with content on a touch-enabled device. We define the term “semantic weight” to categorize perceptions of weight and size in a drag and drop task. We refer to a visual (i.e. text or image) representation of an object as semantically heavy if the object represents a large and heavy real world item (e.g. a car); we refer to a visual representation of an object as semantically light if the object represents a small and light real world item (e.g. an apple). In order to reduce biases as far as possible and keep the size of the study manageable we decided to investigate one specific HCI task (i.e. drag and drop) and perceptions of weight and size using text and image as representation modalities. Based on neuroscience expertise, weight and size were chosen and combined since these features often have a direct correlation in real life. These also characterize how easily or hard an item is to move; semantic weight is a user’s implicit estimation of how easy or hard it is to move a represented object.
Our main hypothesis was that the semantic weight of an object would affect performance, i.e. the time needed to drag and drop the object. We believed that semantically heavier objects would be dragged faster since more effort would be “expected”. In the following, we examine the assumption that only spatial measures influence user performance; we argue that semantic qualities of representations are also significant for user performance and behavior. We present exemplary evidence for a non-spatial factor (i.e. semantic weight) that influences users’ performance in a typical HCI task on a touch-enabled device. Furthermore, we discuss the relation between semantic weight and affordances of the qualities of objects that allow a user to perform an action [6]. Our study results add to a better understanding of potential constraints that vision imposes on tangible interaction techniques and indicate that meaning in textual and graphical representations significantly influences user behavior during tangible interaction.

BACKGROUND

Related Work in Neuroscience

In neuroscience, it was already identified that the meaning of a word (e.g. “large” vs. “small”) affected the way people grab same sized objects. For example, Gentilucci et al. [5] describe studies conducted to investigate if a word (e.g. small, large) printed on an object (i.e. a piece of wood), could affect the movement directed towards that object. In a study with Italian participants, they used the words “GRANDE” (“large”) and “PICCOLO” (“small”); their results showed that the word “large” led to a larger maximum grip aperture than the word “small”. Glover et al. [7] investigated whether words that are only implicitly related to size (e.g. “APPLE” versus “GRAPE”) had an effect on motoric behavior and were able to find that these kind of objects activated motor tendencies related to their features. They also observed that action affordances can be activated by non-target objects that are located in the visual field as well as by word labels attached to target objects [7].

Benoit et al. [2] describe a study that aimed to find out whether performance in visual search tasks is influenced by the real world size of an item. They studied a search task where images of larger and smaller objects were presented on a screen. Participants’ task was, for example, to name the items that were presented larger on the screen. When real world larger objects were presented by larger items and real world smaller objects were presented by smaller items reaction times decreased. When real world larger objects were presented by smaller items (and vice versa) reaction times increased.

Vision is central to perception and action. However, the serial view that action is controlled by what is perceived has often been criticized. For example, Goodale and Milner [8, 18] distinguish between vision for action and vision for perception. Based on evidence from several experiments, they provide a model for two parallel working visual streams, one concerned with action and the other responsible for conscious perception. Consequently, action is not determined by what is perceived alone, but rather the two streams work closely together and contribute to the action.

We agree that perception represents our visual experience of the world, but not that it provides the direct foundation for action. [19]

This means that action directed to a digital representation of an object is not guided solely by the pixels perceived but also depends on information (e.g. object characteristics) derived from previous encounters with the object. Note that seeing an object and perceiving an object is not clearly differentiated. In other words, one might see an object and in that very moment potentially perceive object characteristics, but one might also act on the object characteristics without a deeper and conscious perception of these characteristics. This is more pertinent with nonspatial characteristics (e.g. weight, fragility, texture, surface friction and temperature) than it is for spatial properties (e.g. size, distance, shape and orientation) [7].

Related Work in HCI

Prior work based on Fitts’ law has shown that certain spatial factors (i.e. distance, direction and target size) predict the time needed to drag and drop an object. Fitts’ law is a well known theoretical tool in interface research that describes speed/accuracy trade offs in aimed movements [3]. It predicts that the time needed to point to a target depends highly on target size and distance to the target. McKenzie and Buxton [16] extended Fitts’ law to two-dimensional tasks taking into account confounding variables such as approach angle and target shape. In this paper, we present that the semantic weight is a semantic quality that has, in addition to spatial properties of a drag and drop task (i.e. distance, target size and approach angle), an effect on user performance and behavior (i.e. the time that is needed to drag and drop an object).

Related work in mouse-based interaction exists, in terms of how to analyze mouse movement trajectories. For example, Hurst et al. [12] studied mouse based pointing in order to automatically recognize performance and make computers more accessible to a wide range of users. Gajos et al. [4] provide methods to develop classifiers (e.g. a set of features that explain different kinds of mouse movements) for mouse based interaction and suggest that they can serve as a blueprint for other interactions (e.g. touch screens). We reviewed the work on mouse movement in order to identify a subset of their suggested features that, we believe, fits our goals for analyzing finger movement during a drag and drop task on a touch screen (e.g. mean, standard deviation, and range in finger movement time during the task and Fitts Index of Difficulty).

Research in touch-based interaction aims to understand subtle and systematic differences in touch behavior. For example, Henze et al. [9] and Holz et al. [10] studied target acquisition in touch based interaction. Holz et al. [10] found evidence in a series of three lab-based user studies where
participants had to target objects in a seated position, that the features users align with during target acquisition are visual features, and that these features are located on the top of the user’s fingers, not at the bottom, as assumed by traditional devices. Henze et al. [9] analyzed the touch behavior of smartphone users by publishing a game to the Android Market and collecting more than 120 million touch events. During the game, multiple targets (i.e., circles in different colors) in different sizes were presented to the player and the task of the players was to touch these targets. They showed that touch positions are systematically skewed toward a position in the lower-right of the touch screen. Karlson et al. [13] present a thumb-based interaction technique that improves the interaction with smaller targets on mobile phones. Brandon et al. [22] studied how the way people hold and manipulate objects (i.e., grasp-recognition) can be used as user interfaces. Baudisch et al. [1] studied pointing input capabilities for very small screen devices. Novel user interfaces for behind, under or around device interactions [24, 21, 25, 15] have also been a major topic in the last few years. Those related work examine the benefits of different touch-based interaction techniques focussing on the traditional spatial measures while leaving out the influence of the content that users will interact with, which we argue potentially influences the benefits of different touch-based interaction techniques.

Related work in tangible interaction exists that focuses on the shape of interactive objects. For example, Ware et al. [23] studied factors that may cause a discrepancy when using haptic and visual objects. One of the factors was shape mismatch. In their scenario, a real 3D object was manipulated by the user to modify a virtual object. They studied if shape mismatch (e.g., object held unlike object seen) had an effect on performance. Their hypothesis was that shape mismatch would cause cognitive problems affecting performance. However, they found no significant results to back up their hypothesis. The work of Ware et al. [23] is related to our work by means of a different perceptual channel (i.e., tactile) and semantic category (i.e., shape).

Overall we are interested in understanding differences in touch behavior that is caused by the content and the meaning of the objects that users “see” and interact with. Our definition of semantic weight categorizes some aspects related to what Gibson [6] describes as affordance, the association of characteristics based on prior knowledge from the real world with perceived objects; for example, seeing an apple, a person knows based on prior knowledge that the apple can be grabbed and moved around. However in comparison to Gibson we are interested in the details of the action that is performed based on seeing and perceiving a “known” object. In order to gain a better understanding of how and what semantic factors that are “perceived” visually influence touch behavior, empirical research is needed.

THE STUDY
Our main hypothesis was that the semantic weight of an object would effect the time needed to drag and drop an object. However, we also studied the manifestation of semantic weight in other features related to finger movement during a drag and drop task (e.g. mean, standard deviation, and range in finger movement time during the drag and drop task).

Participants
Twenty-four participants took part in the experiment and received a 15 Euro coupon in return for their participation. All recruited participants were students; most of them were studying law, native German speakers, right handed, and between the age of 20-35 (mean = 25.5, SD = 4.2). The participants reported that they are not colorblind. Furthermore, they considered themselves as experienced (e.g. owned a smartphone or tablet) in using touch-based interfaces. 12 participants were female and 12 were male.

![Figure 1. Abstract presentation of the drag-and-drop task that participants were asked to complete.](image)

Material and Setup
A first generation iPad was used in the experiment. Pixel density on the iPad was 132 pixels per inch. All tasks were performed on the iPad in a seated position within a laboratory environment. The posture of the participants and the position of the iPad were kept fixed as much as possible. Participants could adjust their chair to get to a comfortable sitting and viewing position. The iPad position was fixed on the desk with glueing strips. Participants were instructed to keep their non-dominant hand on the table next to the iPad during the tasks. Three cameras were positioned in the laboratory to record overall posture of the participants, their facial expressions and the iPad display during the task. The duration of the experiment depended on the performance of the participant; however, all participants completed the experiment within 20 minutes. All video inputs were merged into one video stream and routed to an observation area, where participants’ behavior could be observed remotely.

Since we wanted to investigate a predicted effect that is subtle with a task that can be performed within a few seconds we decided to use a study setup for repeated measurement. Doing this allowed us to collect multiple data sets from single participants and work with differences within participants and not between participants. We iterated our task design a few times before deciding to use the final design that is presented in Figure 1. The task of the participants was to drag and drop an object inside an area. Draggable objects were either white circles with text labels on top or images of real world objects.

In the following, we briefly describe the iterations (see also Figure 2). In our first design iteration for the drag and drop

![Figure 2](image)
In order to minimize word length related effects, we chose objects with names that were similar in length for the text modality (i.e. words consisting of four to five letters). Due to the fact that all participants were native German speakers the words were represented in German. The following words were used in the experiment for the text modality:

- Semantically light: Apfel (apple), Ball (ball), Erbse (pea), Nuss (nut)
- Semantically heavy: Auto (car), Mond (moon), Haus (house), Berg (Mountain)

In figure 3, we present the set of images that we used for the image modality.
Our goal for the image modality was to use images that had a square or round form (e.g. a car or a truck photographed from an angle that would end up in a squarish image). This provided typical draggable objects (e.g. app icons) for touch displays and reduced unintended influences due to variations of shapes. Since we were interested in the influence of the semantic of the object, we choose objects that we considered as easily recognizable as real world objects.

**RESULTS**

In order to conduct an analysis we computed measures depicting detailed information on finger movement. The following measures were computed:

- Finger movement time in milliseconds - the time from tapping the draggable object to successfully dragging and dropping the object at the target area
- ID (Fitts Index of Difficulty) - the logarithm term in the formulation of Fitts’s law

We used the Shannon formulation of Fitts law to compute ID (Fitts Index of Difficulty):

\[ MT = a + b \cdot \log_2 \left( \frac{2A}{W} + 1 \right) \]

Based on the distances and target sizes, we used for our task, three values of Fitts Index of Difficulty (ID) were computed. We labeled these three values of ID as “low”, “in-between” and “high”.

**Results on finger movement**

For both presentation modalities (i.e. text and image), we conducted a 2 (distance) × 2 (direction) × 2 (semantic weight) × 2 (target size) × 4 (time) repeated measures ANOVA to compare the effect of semantic weight of the draggable objects on finger movement time. We used a repeated measures ANOVA, since the dependent variable finger movement time was measured under a number of different conditions (e.g. different levels of semantic weight, distance, direction and target size) and the measurement of the dependent variable was repeated 4 times. The distribution of movement times was inspected and showed a marked positive skew typical of response times. We log-transformed the movement times, which restored a Gaussian distribution which is a requirement for a within subjects ANOVA.

**Effect of semantic weight**

From Fitts’ law, we already know that longer distances cause longer movement times; bigger target areas cause shorter movement times. In addition to the expected main effect of distance and target size, we found that semantic weight had a main effect on finger movement time (see Figure 4).

The repeated measurement ANOVA produced the following results. For the text modality, semantic weight had an effect on finger movement time, \( F(1, 23) = 7.2, p < 0.01 \). For the image modality, semantic weight had an effect on finger movement time, \( F(1, 23) = 24.7, p < 0.001 \). The results show that, in both representation modalities, semantically heavier objects (e.g. a car, a truck, a mountain or a house) were dragged significantly faster than semantically lighter objects (e.g. an apple, a ball, a pea, a nut). Thus the action tendency activated by semantically heavier objects led to a faster dragging of the heavier objects. When objects were represented as images the mean movement time for semantically light objects was 771 ms (SD = 361 ms) and 728 ms (SD = 365 ms) for semantically heavy objects. When objects were represented as text, the mean movement time for semantically light objects was 832 ms (SD = 457 ms) and 741 (SD = 379 ms) for semantically heavy objects.

The independent variable direction interacted with semantic weight in both modalities. The effect of semantic weight appeared dominantly when participants had to drag from right to left. Going through the video records helped us to realize that the drop area could have been occluded or partly occluded when dragging from left to right, depending on participants’ hand and finger position. We believe that this could be part of the reason for losing the attention on the dragged object and focusing on finding the target.

However, since we tested our design beforehand, we believe this is unlikely the sole reason. We do not have a final explanation for this interaction; however, since all our participants were right handed, we believe that the effect has to be related to the nature of the task.

**Effect of the representation modality**

In our study, we used two different visual representations and designs for draggable objects. We represented objects as text and as images. Although the differences in mean values are quite similar for both modalities (see Figure 4), our statistical analysis shows that the effect was higher for the image modality. In Figure 5, we present for both modalities and for each of the 24 participants the delay in milliseconds that semantically heavier objects caused. Figure 5 also explains why the statistical effect was higher for the image modality, while the difference in finger movement time for the
mean values over all participants is quite similar. More participants dragged, on average, semantically heavier objects faster in image modality than in text modality (see also the number of positive and negative values for both modalities in Figure 5). Although some participants dragged semantically heavier object “much” faster then semantically lighter objects in text modality (e.g. participant 1, 3, 4 and 24), more participants were effected by semantic weight for the image modality (e.g. only participant 17 dragged semantically lighter objects faster than semantically heavier objects in image modality).

Figure 5. The mean difference in finger movement time (time for semantically light objects - time semantically heavy objects) that is caused by semantic weight for each participant (both modalities).

Our main intention was not to empirically compare the effect of the representation modalities, but to explore the effect of semantic weight in finger movement for two separate designs based on different representation modalities. However, our study setup allowed us to combine the data we collected for both modalities and add the representation modality as an additional independent variable. The analysis on this combined data further supports our hypothesis that semantic weight has a significant effect on finger movement during a typical drag and drop task on a state-of-the-art device. With semantic weight, a non spatial measure we identified, has a significant effect on finger movement time (F(1, 23) = 2.4, p=0.13).

We did a separate analysis taking ID (Fitts Index of Difficulty) into consideration. For the image modality (F(4, 138) = 4.98, p=0.025) and text modality (F(4, 138) = 31, p<0.001), ID interacted with semantic weight. The interaction was also apparent on the combined data (i.e. data from both modalities) F(4, 138) = 31, p<0.001, see Figure 6.

The expected effect of increasing ID is that task completion times increase. We observed this expected effect also for tasks with semantically light and heavy objects where the mean movement times increased with increasing ID. However, the difference of mean movement time between tasks with semantically heavy and semantically light objects is big when the ID is low and when the ID is high. For the in between value of ID there seems to be no difference of mean movement times between semantically heavy and light objects.

The high value of ID is related to the tasks where distance is long and target size is small. The low value of ID is related to the tasks where distance is short and target size is large, while the “in between” value of ID is related to a set of tasks where distances and target sizes are mixed (i.e. short distance and small target, long distance and large target size).

DISCUSSION

One explanation for the ID-related effect, presented in Figure 6, is based on our “naive” combination of real world size with real world weight. We assume that dropping a representation of a real world large object into a smaller target area caused an additional effect that we did not predict. Consequently one could hypothesize that a formula like this: \( \log_{2}(\frac{2A + \text{someValue for SemanticWeightOfObject} + 1)}{W + \text{someValue for SemanticSizeOfTarget}}) \) would be one that takes into account the “difficulties” caused by the semantic qualities “weight” and “size”. We leave research on this hypothesis for our future work.

In order to be sure that our results were not caused by a chosen single represented heavy or light object and representation that we chose we checked the mean values for the single objects. For example, since for the image modality the strawberry image seemed to produce lower mean movement times than the other images (see Figure 7); we checked the values and made sure that the main effect of semantic weight was not caused by this single object.

In this paper, we successfully demonstrated that semantic weight, a non spatial measure we identified, has a significant effect on finger movement during a typical drag and drop task on a state-of-the-art device. With semantic weight, we categorized object characteristics that are based on prior knowledge from the real world. Semantic weight is one categorization that fit the kind of direct manipulation task we were interested in (i.e. drag and drop task). In general, the object characteristics that we refer to are similar to affordances, potentially endless, and should not be simplified.

In addition, HCI relevant tasks are usually performed under dynamic and complex circumstances that may or may
not amplify the influence of some semantic qualities (that are potentially hidden from users and designers). Hornecker [11] scrutinizes the assumption that users’ prior knowledge from the real world can be seamlessly transferred to tangible interfaces.

The affordances of physical objects are potentially endless and users creatively select those that fit their understanding of the system, their aims and the situation. Designers capability to design affordances into objects, let alone restrict them to desired ones is thus limited [11].

We also argue that one has to be very careful in defining object characteristics as non-spatial measures since object characteristics are indeed manifold. Related work in neuroscience provides us with examples (e.g. weight, fragility, texture, surface friction and temperature [7]); however, it is our responsibility to transfer those to HCI relevant tasks and designs, and to explore new object characteristics that matter using the tools that we, as HCI researchers, have at our disposal (e.g. explorative design approaches, qualitative research and research in the large).

In tactile and tangible interfaces, vision is constantly used for perception and to guide a user’s actions. We believe that the traditional notion of a serial view on perception and action is widely distributed in the HCI community. Goodale and Milner [8, 18] criticize the serial view on vision and action; they provide an alternative model on the visual system which provides new insights on how vision works in detail. These insights can help understand potential constraints vision imposes on tangible interaction techniques which is becoming more pertinent than ever with the growing number of natural, direct and movement-based interaction paradigms and technology improvements capable of capturing real-time and subtle movements.

Milner and Goodale [19] also put emphasis on the importance of the task and that even when performing an apparently simple task (e.g. pinching, zooming and flicking), one cannot help but simultaneously perceive the objects and, often, also the hand reaching out towards it.

It is a perennial problem in psychology that no one task ever provides a pure measure of any given mental or neural process. The brain, and indeed our behaviour, are far too complex for that ever to be possible. [19]

In order to uncover new parts of the story behind what makes an interface design in its variety cause different behavior based on vision and perception, we need to tackle the measures we use to analyze and explain user behavior.

The ability to analyze finger and hand movement is necessary to answer research questions and the traditional spatial measures have proven to be valuable and will do so in the future. However, with upcoming potentials to collect and analyze large sets of data from tactile and tangible interactions (e.g. Henze et al. [9] collected more than 120 million touch events), new possibilities arise for varying experimental tasks and to adjust the measures we use for our analysis. We argue that one should use non-spatial measures (i.e. semantic qualities) in addition to the established measures for analyzing user behavior that is based on vision.

CONCLUSION
In our first attempt to research the influence of semantic qualities on finger movement and user performance, we used a lab-based study and repeated measurements. We reproduced predicted effects from neuroscience (i.e. the meaning of words and images activate action tendencies related to the meaning of the represented objects) and demonstrated in a touch-related typical HCI task their manifestation in finger movement on a state of the art touch-enabled device (i.e. semantically heavier representations were dragged significantly faster than semantically lighter representations in a drag and drop task).

While we studied touch-based interaction our results should be relevant for tangible interaction. As some of the related work in neuroscience that we presented in this paper investigate how vision and semantics influences grasping behavior, we believe that our results will transfer to tasks with physical tokens. Our results indicated that users will complete tasks quicker if semantically heavier representations (e.g. icons, labels) are used. So far, we studied semantic weight in isolation. In our future work, we intend to investigate the influence of semantic qualities in realistic task contexts.

ACKNOWLEDGEMENTS
The financial support by the Federal Ministry of Economy, Family and Youth, and the National Foundation for Research, Technology and Development is gratefully acknowledged (Christian Doppler Laboratory for “Contextual Interfaces”). Furthermore, we would like to thank Elisabeth Stöttinger for her valuable help regarding the related literature in neuroscience.

REFERENCES


