MASTER

Effect or demand
the influence of effortful interaction on visiual perception

Hamelink, C.J.

Award date:
2014

Eindhoven, January 2014

Effect or Demand: The Influence of Effortful Interaction on Visual Perception

C.J. Hamelink

Identity Number: 0740869

In partial fulfillment of the requirements for the degree of

Master of Science in Human-Technology Interaction

Supervisors:
prof. dr. W.A. IJsselsteijn Eindhoven University of Technology, department of IE&IS
dr. D. Lakens Eindhoven University of Technology, department of IE&IS

Keywords: Computer Mediated Environment, Effort Theory, Effortful interaction, Visual perception, Weight

Abstract

Recent developments in the field of human-computer interaction demonstrate that embodied interaction is becoming more common. This type of interaction tends to be seen as fun and natural but has one important downside, its reliance on bodily movement is effortful and can cause user fatigue (e.g. Cabral, Morimoto & Zuffo, 2005). While effort in interaction is known to exist, up until recently few researchers have considered the consequences of effort for the user beyond user comfort. Recently, however, such an influence was proposed. Van Beurden (2013) considered the effect of interaction gain for various input devices (e.g. mouse) and found that at low settings, and thus at larger bodily movements, distance reports for objects in mediated environments were higher for constant presentations. Based on the notion of Effort Theory it was argued that an effort mechanism may underlie these findings. This notion assumes that experienced effort may alter one's visual perception (Proffitt, Bhalla, Gossweiler & Midgett, 1995; Bhalla & Proffitt, 1999). Currently, the scientific debate on this theory is highly topical and due to mixed empirical findings it is controversial. In light of the Effort Theory discussion, the suggested Effort Theory mechanism in human-computer interaction deserves further empirical clarification.

The goal of the current research was to determine whether effortful human-computer interaction influenced visual perception and user experience. Based on a consideration of literature and pilot study results it was hypothesized that (1) effort effects of visual perception on human computer interaction would be found; (2) that inducing effort would deteriorate pragmatic user experience; and (3) that inducing weight would lead to an increase in hedonic user experience.

Two experiments, based on Experiments 8 and 9 by van Beurden (2013), were conducted to investigate these hypotheses. In the first experiment participants estimated distances between two sequentially presented buttons on a computer display and effort was induced by manipulating mouse weight. After the experiment, experimental suspicion and user experience were assessed. Users were unaware of the experimental goals and no distance estimate differences between weight groups were found. In line with the hypotheses, however, there were indications that user experience may be dependent on effort. The second experiment consisted of a similar distance estimation task in a Virtual Environment. In this experiment, a stronger weight manipulation using a weighted 3D tracking device was applied. As with the first experiment, participants were found to be naïve about the experimental intentions and no effect of effort on distance perception was found. The user experience findings, however, surprisingly deviated from the hypotheses.

Based on the findings from both experiments it can be concluded that Effort Theory does not reliably predict the results in either experiment. While these results could be seen as problematic for Effort Theory, an alternative explanation could be that the manipulation was too weak to cause an effect. Despite the fact that comparing exerted effort between conditions with van Beurden (2013) is difficult, the current work indicates that it is unlikely that earlier gain based findings can be explained in full by relying solely on an effort mechanism. The mixed user experience findings make it difficult to draw firm conclusions but have led to new and interesting questions about the interplay between experience of a device and the interaction user experience. Overall, the current work could be seen as one of the first to have investigated effortful interaction on visual perception and user experience. Future research could accommodate for identified limitations and demonstrate whether effort should be seen as an influential variable in human computer interaction.

Preface

Here before you lies the fruit of several months of hard work. While working on this project I have many times scrolled past an empty preface section occasionally imagining what it would be like writing it. Now that the time of actually writing it has come, I am experiencing mixed feelings. Both feelings of satisfaction because I was able to pull this off, and feelings of sadness as its completion marks the end a great era. What the future holds I am not sure, but if it is anything like the time I have had at the university then I am up for some good times to come.

I started this master track with the intention to learn more about Human-Computer Interaction on a scientific level and I can honestly say that I got everything I wanted and even more. Throughout my master years I have done such varied things as map human behavior and preferences using qualitative research methods, tested human performance and memory, co-developed a Kinect controlled beat-box application, and even attempted to destress people using a stroking machine. Some of these I had never imagined I would do and I probably wouldn't have done them if it weren't for the opportunities provided by the program. I cannot but conclude that I have had a blast during each and every year of it.

For this project there are of course many people that I want to thank. First my supervisors Wijnand IJsselsteijn and Daniel Lakens who have been both a source of inspiration to me as well as a sounding board for my own ideas. Many times after our meetings I have felt fully recharged and ready to go at it again. Second, the university staff that have assisted me with many small but important things such as modifying devices and setting-up my research environments. In particular I should name Aart van der Spank who freed up a lot of time to assist me. Finally, I want to thank my family and both university and hometown friends for their seemingly limitless support, expressed interest in my research, and for the provided distraction when it was necessary. Without the above mentioned people I wouldn't be where I am right now.

Niels Hamelink

Huizen, January 2014

Table of Contents

1. Introduction

Ever since the introduction of the first computer, researchers have investigated novel ways in which humans can interact with these systems and what the consequences of using different interaction styles are. In the previous decade, the motion control based Wii gaming console brought a new type of human-computer interaction to the masses and started a new interaction trend which currently is still ongoing. In an attempt to increase the fun and immersion of playing video games, the Wii controller and later devices (e.g., Microsoft's full body tracking Kinect camera) support an increasing amount of bodily involvement in interaction. For video games this means a movement away from traditional button presses to trigger in-game actions. Instead, the action is to be acted out by the player through bodily movement. Possibly due to the low costs and traction that these devices got in the consumer market, they were also proposed for different, more serious applications. For example, the practicality of hands-free input has led the Kinect interface to be suggested for highly demanding environments such as operating rooms (Ruppert, Reis, Amorim, de Moraes, & da Silva, 2012). Beyond the influence on fun and the practicality in specific scenarios, embodied interaction tends to also be seen as more natural as the interaction is more akin to interaction with the real world (Abowd & Mynatt, 2000). While this feature is often lauded as an advantage, there is one problem with this closer resemblance. The stronger similarity to interaction with real world objects makes embodied interaction more physically effortful to perform. While interaction effort is known to affect user comfort due to the introduction of fatigue (Lyons, Lopez Silva, Moher, Jimenez Pazmino & Slattery, 2013), recently it has been suggested that the consequences of effortful interaction may extend beyond such discomforts and may affect distance perception for stimuli in one's environment (van Beurden, 2013). When considering the plausibility of such a suggestion, a more general question regarding the malleability of visual perception needs to be asked first. More specifically, it should be considered whether physical effort can influence human visual perception.

Whilst perhaps not realizing it, many may hold anecdotal experiences that could lead to intuitive support of an influence of effort on distance perception. Take the example of having been on a long walk. When the end destination approaches, it may seem like it is farther away than it actually is. Considering the factors that change during the walking activity, physical potential may be one of the first that comes to mind. Indeed, acting upon the world costs energy so after a long walk, one has spent energy and may feel physically tired. Tiredness then forms a description of one's current bodily state. Based on this reasoning, a seemingly obvious question would thus be whether these bodily states influence perception. Interestingly, despite the apparent intuitiveness of the mechanism, scientific interest in the role of action on perception was sparked only a relatively short time ago following the introduction of a novel view on perception (Gibson, 1986). This work inspired others to investigate the influence of embodied factors on perception. Throughout the last decade this line of work ultimately led to the introduction of the Effort Theory[1], which argues in favor of an

[1] The broad variety in methodologies and experimental tasks used in the experiments have led the theory that stemmed from them to come to be known under different names such as 'action-specific'- or 'embodied' perception, or 'effort theory'. Because all terms describe the same fundamental idea namely the relation between bodily states and perception, but the focus of the current thesis lies on effort, the term Effort Theory will be used.

influence of energetic costs on visual perception (Proffitt, 2006). Despite the intuitiveness of the account and empirical evidence that has been put forward, the next chapter will show that the theory is heavily debated and even disputed by some.

Returning to the main scope of Human Computer Interaction, the before mentioned study by van Beurden (2013) investigated the influence of interaction gain or amplitude of arm movement on distance perception. It was found that distance estimates were higher when arm movements became larger. One of the mechanisms that was coined to explain the effect related to the additional effort required in low gain interaction. In order to determine whether these findings can indeed be attributed to an effort-based account, they should be considered within the context of the Effort Theory discussion and literature. Hence, the following chapter will focus on three main points. First, it will provide an overview of the development of visual perception and Effort Theory, and will discuss the current state of the Effort Theory. Second, it will discuss the characteristics of interaction with Computer Mediated Environments. Finally, it will connect the dots by analyzing indications of a possible influence of interaction induced effort on distance perception taking into account what is known about Effort Theory. Following this discussion, the hypotheses of the current thesis will be outlined.

2. Related work

2.1 The Evolution of Views on Visual Perception

Traditionally, visual perception has often been studied from a constructivist perspective. Constructivists assume that incoming visual information is insufficient and requires intermediate processes to allow humans to specify aspects of the world (Gordon, 2004). Thus, perception is thought of as indirect in nature, utilizing memory and stored schemata to interpret incoming sensory information. This assumption has had a strong influence on the types of perception experiments that have been conducted in the past. Constructivists have mostly focused on studying the mechanisms that underlie perception by using cue reduction methods that aimed to selectively control information (or cue) exposure, thereby frequently depriving participants from perceptual cues in experimental settings. Examples of such cue reduction approaches include only briefly exposing participants to stimuli or have participants perceive stimuli through peepholes (Norman, 2002).

Dissatisfaction with this classic view led Gibson (1986) to propose an alternative perspective which would come to be known as the ecological view. This view rejected many of the constructivist principles and methods including its focus on determining the underlying principles of perception. The ecological view assumes that perceivers are active and therefore argues that the static reduction based methods do not suffice for studying perception. A key concept within the ecological view is the concept of *'affordances'*, a term coined by Gibson to denote "*what* [objects or environments] *afford for the observer*" (Gibson, 1986). The concept of affordance acknowledges that perception is in essence a relationship between an active organism (with its own unique perceptual properties) and its environment. Affordances are highly relevant in directing action due to their information on behavioral potential (Greeno, 1994). An object can have multiple affordances such as 'eat' and 'throw' and thereby give rise to many actions. An important assumption is that affordances tend to be immediately perceived when an environment or object is looked at (Gibson, 1986). Thus, unlike the classic view, the Ecological account does not assume that a physical object acquires meaning but rather that perception is direct. Gibson's ecological account raised a scientific discussion which led to criticism on the view. While some of the early objections did not seem problematic for the account (for a discussion see Michaels & Carello, 1981), the discussion on Ecological perception and developments of the account is ongoing and it seems that some conceptual challenges may still exist (Michaels, 2003).

Despite the criticism on Ecological perception, the foundation of this view remains relevant as some of its key assumptions have been incorporated in other theories. For example, the rejection of an essential role for amodal objects resonates with the ideas within grounded cognition (for a review see Barsalou, 2008). This holistic theory assumes that cognition is grounded in simulation, situated action and bodily sates. Ultimately, Gibson's ideas also inspired early studies that investigated a potential role for effort in visual perception.

2.2 Effort Theory

2.2.1 An Influence of Effort on Visual Perception?

Two studies have laid the foundation for the idea that effort could influence visual perception. In a pivotal but controversial study, Proffitt, Bhalla, Gossweiler and Midgett (1995) researched prior claims that suggested a tendency towards overestimation of geographical hill slant. Four experiments were conducted to consider slant estimates from both the top and bottom of a hill, and a fifth study researched the influence of fatigue on such estimates. It was found that hill slants judged from the top were estimated as higher than when judged from the bottom, and that hill slant estimates increased with fatigue. Following these results, an additional series of hill slant estimation experiments was conducted using similar methods (Bhalla & Proffitt, 1999). Comparisons were made between heavy backpack and non-backpack equipped; pre- and post-run; fit and non-fit and elderly and young participants. Across all experiments the same pattern emerged, verbally reported hill slant estimates were higher for groups that were expected to have to exert more effort, or were less physically capable. Specifically, this applied to backpack equipped, post-run, non-fit and elderly participants. The findings from both studies appear to be compatible with the preexisting ideas of affordances (Proffitt et al., 1995) but go beyond the ideas expressed in the ecological view as they suggest that depending on one's physiological state, a constant hill slant is perceived differently. In other words, perception should be thought of as influenced by capabilities, bodily states and goals (Proffitt, 2006; Witt, 2011; Proffitt, 2013), a notion which would become known as Effort Theory.

Over the last decade, the encouraging results from the formative studies have led scientific interest in the Effort Theory to flourish. During this time, both partial replications of the hill slant studies, as well as entirely different methods have been employed to test the account. Similar hill slant differences as in the foundational studies were also reported following walking effort manipulations in VR (Creem-Regehr, Gooch, Sahm & Thompson, 2004) and inverse differences were found with increasing blood-glucose levels (Schnall, Zadra & Proffitt, 2010). The account was also reported to generalize beyond hill slant to different forms of perception. Manipulations of walking effort (Proffitt, Stefanucci, Banton & Epstein, 2003) and ball throwing effort (Witt, Proffitt & Epstein, 2004) revealed that perceived distance was similarly affected by effort inductions; a series of athletic performance related studies indicated that with higher performance, estimations of ball size (Witt & Proffitt, 2005) and goal size (Witt & Dorsch, 2009) increased. However, not all studies were able to replicate these effects (Hutchison & Loomis, 2006a; Woods, Philbeck & Danoff, 2009). These failed replications formed the start of a series of studies that scrutinized past confirmatory findings which resulted in alternative explanations for the effect. Several studies provided evidence for the involvement of experimental demand in the results (Durgin, Baird, Greenburg, Russel, Shaughnessy & Waymouth, 2009; Durgin, Klein, Spiegel, Strawser & Williams, 2012), and other alternative explanations have also been coined for some of the findings (Cooper, Sterling, Bacon & Bridgeman, 2012; Firestone, 2013).

2.2.2 Topical Disputes in Effort Theory

Currently, Effort Theory remains a highly controversial theory and there are reasons to believe that the debate may not easily be resolved. Considering the criticism on Effort Theory cited in the previous section, it should be noted that there is an apparent lack of consensus on visual perception beliefs such as whether perception from memory can be seen as perception. Awareness of the different beliefs regarding visual perception is very important as any study that attempts to test the Effort Theory is likely to be confronted with this issue. The lack of agreement within the field has consequences for the discussion on which measurement of visual perception constitutes an unbiased measurement. Typically, a distinction is made between direct and indirect measurements. Many of the early studies used direct measurements in the form of verbal reports. Such measurements directly assess the variable of interest (e.g., perceived distance) thereby possibly informing participants of the experimental goal and making reports more vulnerable to post-perceptual bias (Gogel, 1976). As an alternative, several indirect measures have been coined. Examples for egocentric distance (the distance between viewer and a point in the world) include the size-distance coupling, blindwalking and blindpulling (Loomis & Philbeck, 2008; Philbeck, Woods, Kontra & Zdenkova, 2010). These measurements require the participant to generate a response that cannot directly be related to the variable of interest.

Given that the sensitivity of direct measurements has been known for a long time, the choice for verbal reports may seem remarkable. However there are various reasons as to why one would not opt for an indirect measurement. The most important of which concerns a less frequently made distinction that can be made between measurements namely the time at which the response is given. Several of the indirect measurements occur after a stimulus has been perceived (for an example see Table 1). In the past, some have argued that any response outside of full sight is based on memory and therefore a measurement of memory rather than perception (Proffitt et al., 2006; Cooper et al., 2012). While not all Effort Theory researchers subscribe to this view (e.g., Hutchison & Loomis, 2006b), the lack of consensus on this aspect shows a significant difference in views on visual perception.

	Outside of perception	**Under full perception**
Direct	-	Verbal reports[2]
Indirect	Blindwalking; Blindpulling	Size-Distance coupling[3]

Table 1 Non-exhaustive list of egocentric distance measures mapped on both directness and time of response

A logical conclusion that could be drawn from this discussion is that the ideal response is both indirect and provided under full perception. Based on Table 1, one could argue that for egocentric distance estimates, the size-distance couplings seem ideal as they meet both criteria. However, the validity of underlying assumptions of this measure has been disputed (Haber & Levin, 2001. Because finding a measure that meets all criteria is complicated, often suboptimal alternatives will have to be selected which provide grounds for further discussion. One approach to accommodate for this problem would be to use several measurements and rely on the principle of convergence (Loomis & Philbeck, 2008), however, some have warned for an overreliance on such an approach (Woods et al., 2009). Judging from these observations, in the near future, Effort Theory's status as a controversial notion may continue to last.

[2] Verbal reports can theoretically occur outside of full perception. However, most studies appear to have used this measurement in full perception.

[3] Size-distance measurements have also been used in combination with impaired perception procedures however as these procedures do not involve memory they are characterized as 'under full perception'

2.3 Interaction with Mediated Environments

Interaction between humans and computers occurs through Mediated Environments (ME). Two types of mediated environments can generally be distinguished. High immersion environments that often rely upon a Head Mounted Display and sometimes a plethora of trackers and sensors, known as Virtual Environments (VE), and environments that generally feature a lower level of immersion and are displayed on a computer screen, also known as desktop environments (DE). Comparing MEs to the real world shows that these environments have the potential to provide superior control over what an environment should look like and how it should function. Despite their flexibility, however, MEs are bound by one rule that also applies in the real world. That is, to make alterations, interaction with the environment is required. Like interaction with the real world, for most of the currently available interaction methods this means that some amount of physical effort is required to alter a ME. Non-immersive setups such as desktop computers have traditionally often been interacted with using a computer mouse. While computer mice are known to be able to cause strain and fatigue under some circumstances (Wahlström, 2001), usage of this device generally does not require much effort. Hence, the amount of effort required to manipulate MEs has traditionally been low. However, recently both types of ME have been subject to a trend that may decrease the effort gap between the real world and MEs. As indicated in the introduction, low cost consumer devices such as the Nintendo Wii and Microsoft Kinect have led to broad adoption of motion controlled interfaces in a wide variety of environments. The increasing involvement of the body has led this type of interaction to be referred to as embodied interaction[4].

Researchers and designers seem to have generally recognized that embodied interaction can require more effort to perform but while this characteristic may appear to be inherently undesirable, interestingly it has also been used in meaningful ways. Video games developed for the before mentioned Wii and Kinect consumer devices have been used as a means of exercise and are also known as *exergames*. Research has shown that the repeated and occasionally intensive bodily motion required to play exergames can lead to energy expenditure similar to light- to moderate-intensity physical activity (Peng, Lin & Crouse, 2011) and should be sufficient to improve and maintain health (Hurkmans, Ribbers, Stereur-Kranenbeurg, Stam & van den Berg-Emons, 2011). Another way in which effortful interaction has been used in a meaningful way is in learning and formation of memory. Lyons, Slattery, Jimenez Pazmino, and Moher (2012) proposed a series of guidelines for employing effort in interaction so that it can support learning the consequences of an action or inaction. An interesting example application using this principle is also provided by the same researchers. Through effortful embodied interaction, this application allows people to experience the consequences of climate change on the amount of effort polar bears have to exert to travel. A later study showed that this application successfully communicated the message intended for players of the game (Slattery, Lyons, Lopez Silva & Jimenez Pazmino, 2013).

The above examples contain an effortful interaction component that is purposefully implemented and evoked by choices made in interaction design. However, it appears that even when embodied interaction is designed to be used less intensively and has a brief duration, these

[4] Whereas some appear to hold a more strict definition of embodied interaction (e.g Dourish, 2004), recently the term has been used in a more lenient way to also include bodily motion based Interaction (e.g. Lyons et al., 2012; van Beurden, 2013). The current thesis will follow these recent studies and will also use the embodied interaction terminology to describe motion based interaction.

interaction styles can be experienced as more effortful. Several researchers have reported higher levels of fatigue in comparison with traditional input devices for simple performance related tasks such as navigating through a three dimensional environment (Cabral, Morimoto & Zuffo, 2005; Yoo, Han, Choi, Yi, Suh, Park & Kim, 2010). Because in task performance environments, reduced user comfort stemming from the additional required effort will often be unrelated to the task at hand, it is mostly seen as is undesirable. Researchers have therefore been concerned with trying to determine the causes of fatigue and forming design guidelines to inhibit its occurrence (e.g. Baudel & Beaudouin-Lafon, 1993). While such guidelines may help reduce effort required and thus the resulting fatigue, it may be difficult to achieve effort levels comparable with traditional interfaces especially when the interaction is of longer duration. When interaction takes longer and embodied interaction occurs without limb support, it has been suggested that the weight of one's arm can become a load and cause fatigue (Zeng, Hedge & Guimbretiere, 2012), an idea that has seen some experimental support (van Beurden, 2013). While providing limb support seems like a logical solution for this issue, it cannot in all cases be a resolution for interaction fatigue. Depending on the implementation and context of the interaction, finding support may not always be an option. For example, when interaction takes place while standing or when the body is tracked in a more holistic way. Because effort appears to be an inherent factor for some styles of interaction, more knowledge on how it affects users should be attained. Especially given that some of these interaction styles are considered for tasks where performance is of critical importance and errors can be costly such as in surgical procedures (Padoy & Hager, 2010).

2.4 Can Effortful Interaction Influence Distance Perception in MEs?

In light of the increasing applications of embodied interaction interfaces, which appear to be inherently effortful, and what is known from Effort Theory, it seems interesting to consider whether visual perception of MEs is affected by interaction effort. To date, few seem to have researched such a question. Indeed, while MEs have been employed early on to test the Effort Theory notion (e.g. Creem-Regehr et al., 2004), this has mostly been done to replicate experiments conducted in the real world with interaction methods more fitting for this setting (e.g., walking on a treadmill). Types of interaction more common in performance based tasks such as pointing seem not to have been considered. Not only does the literature provide little guidance in answering this new question, in many cases the results of the studies that have used MEs can be called into question either due to methodological limitations (as discussed in section 2.2.1) or due to the immaturity of the employed technology (Durgin & Li, 2010). Looking beyond Effort Theory literature, however, there appear to be grounds for further investigation of an effect of effortful interaction on visual perception.

In a recent study, van Beurden (2013) conducted two studies to investigate whether a mismatch between physical movement and projected movement in both a DE and an immersive VE influenced visual perception by manipulating the gain setting. In the low gain setting, large real world movements had to be made, while in the high gain setting small real world movements would cause the same pointer displacement in the ME. It was hypothesized that larger gains would influence perception such that distances in the DE and VE would be seen as smaller. Indeed, the experiments seemed to show the hypothesized effect both for exocentric (distance between two points in the world) and egocentric (distance between the observer and a point in the world) distance estimations. While the results from this study may sound promising to those who see a role for effort in lower gains, it does not automatically lead to support for the Effort Theory for two reasons, both related to the methodology. First, van Beurden argued that the use of a gain manipulation precludes drawing definitive conclusions as effort is not the only variable that is manipulated. Therefore, two mechanisms accounts in the form of proprioception, or our sense of position and movement, and efferent copy, or outgoing motor commands to the brain, were coined that could also support the findings. Second, even when other explanations could have been ruled out, the results should be subject to the same careful scrutiny that other Effort Theory studies have received. When doing so it can be noted that van Beurden's study has used an interesting combination of procedures that would likely not directly be accepted by Effort Theory critics. Placing van Beurden's measurement approach on a measurement categorization table such as used in section 2.2.2, we can note that the approach is unique as it mostly occurs both outside of direct perception and is direct. Indeed, the results were gathered either through verbal or typed reports which is a direct measurement as these reports form the variable of interest. Moreover, stimuli were not presented under full perception as objects were presented sequentially with the first object disappearing as the second became visible. Sampling from memory was therefore required to make a distance estimation between the two objects.

	Outside of perception	**Under full perception**
Direct	Van Beurden (2013)	Proffitt et al. (2003); Witt et al. (2004); Hutchinson & Loomis (2006); Van Beurden (2013)
Indirect	Witt et al. (2004); Hutchinson & Loomis (2006)	Hutchinson & Loomis (2006)

Table 2 Overview of some Egocentric distance measures used in Effort Theory studies in comparison to van Beurden

While the consequences of measuring perception from memory for the measurement validity remains open for discussion (see section 2.2.2), the directness of the measure without assessment of bias makes it a troublesome method for investigating an Effort account. Therefore, it remains difficult to assess the applicability of an effort-based explanation as an underlying mechanism to explain van Beurden's findings. An adapted replication that limits its manipulation to effort and accounts for common Effort Theory criticism seems desirable.

2.5 The Current Research

The first goal of the current thesis is to consider whether the Effort Theory mechanism functions with interaction with MEs. Despite that the highly focused character of this question, the results of this study have the potential to contribute not only to knowledge on human-computer interaction but also to the Effort Theory literature. Since few studies appear to have manipulated only physical effort in interaction with MEs, an additional goal will be to investigate whether effortful interaction impacts the user experience. Hypotheses regarding both goals will be discussed below.

The Effort Theory hypothesis as considered is based on Woods et al. (2009) who have proposed the effort-calibration hybrid position. This hybrid position assumes that under certain, unspecified conditions, effort can have an effect on perception while in some scenarios response bias can cause misleading results that seem to confirm an effort account. This position was chosen as a starting point as it takes no a priori stance either in favor or against Effort Theory and thereby provides the capability to account for experimental factors such as the environment and methodology in the formulation of the hypothesis.

The current studies will attempt to extend the earlier findings by van Beurden (2013) and therefore adopt the methodologies of experiments 8 and 9. To form a hypothesis for the current work it should therefore be assessed whether effort effects are possible within the context of these methodologies. Considering the ball throwing study by Witt et al. (2004), it can be seen that small weight manipulations have been indicated to be capable of causing an effort effect. The results from this study provide the best indication from Effort Theory literature about the possibility of an effect in a human-computer interaction scenario as the throwing task also concerned effort exertion with the arm. While the ball throwing study suggests that results can be expected, a more important question is whether such an effect would be genuine. When looking at Effort Theory studies that have applied weight manipulations some have done so in a way that seemed unrelated to the task at hand (e.g., providing a backpack before estimating a hill slant). In this sense, the methodological approach in the current study will be somewhat different as the effort inducer is an apparatus that is required for performing the interaction. Since devices are a common way to interact with the ME, no explicit request of using a device has to be made. Because no explicit request is made the device characteristics are expected to be less salient, which in turn may decrease the likeliness of the device being noted as a possible manipulation. This belief was confirmed in a pilot study based on the first experiment which indicated that aspects more closely related to the task (the way in which distances were presented) were more likely seen as the main goal of the experiment. For these reasons, the following hypothesis was formed:

H1. *There is an effect of effortful interaction on distance perception*

The second goal of the current research is to investigate the effect of effort on the user experience. Factors pertaining to the user experience were also explored in the pilot study. Here, a trend was found that seemed to indicate that adding weight increased ratings of perceived quality and durability. Moreover, participants were more likely to think that the apparatus was meant for a specific audience (such as video game players). A limitation of the measurements from this pilot was that the items for quality and durability were ambiguous which makes this measurement noisy. Thus, in order to gain a better insight into whether user experience is impacted by the induction of effort, a reductionist user experience model known as AttrakDiff2 (Hassenzahl, 2007a) will be employed to help specify the type of quality experienced. The model's reductionist nature can be seen in the small number of constructs it relies on to make user experience measurable (Hassenzahl, 2007b). The model distinguishes between pragmatic or task-related quality, and hedonic quality which relates a product to the self. Pragmatic quality and effort are expected to be inversely related, that is, with an increase in effort, pragmatic quality is expected to decrease as executing a task becomes more difficult. Hedonic quality, on the other hand, is expected to increase with effort. The fact that an apparatus was seen as made for a specific audience and more durable when its weight increased is hypothesized to be related to higher Hedonic or 'be'-qualities that were attributed to the apparatus.

H2. *Introducing effort in interaction reduces perceived pragmatic quality*

H3. *Introducing effort in interaction increases perceived hedonic quality*

While it is important to elucidate the aims of the current thesis, in this case it is also relevant to discuss what this thesis is not. While the findings from the current work may contribute to the Effort Theory discussion, it does not primarily aim to prove or reject the Effort Theory as a whole. The goal is therefore not to provide the 'holy grail' in Effort Theory research. Rather, it is an attempt to make informed methodological decisions based on Effort Theory literature to investigate a possible influence of a specific type of effort in a specific context.

3. Experiment 1 – Effect of Mouse Weight on Distance Perception

Recent findings have shown that variations in apparatus gain, the relation between real world movement and computer mediated world movement, influences visual perception (van Beurden, 2013). While several mechanisms have been proposed, the relative contribution of each could not be determined. Following the first hypothesis, the main goal was to determine whether an effort-based account could explain the gain study findings by ruling out other explanations. To achieve this goal, the current study partially adopted parts of the methodology used in van Beurden's Experiment 8. Participants did the same task namely, estimated exocentric distances (distance between two displayed objects) in a mediated environment after having interacted with a computer mouse. However, following a suggestion by van Beurden, the manipulation was changed. Instead of mouse gain, the mouse weight was manipulated such that only the required effort would differ between participants. Additionally, a post-experimental assessment of participant suspicion and user experience in terms of perceived pragmatic and hedonic quality was done.

3.1 Method

3.1.1 Design and Measures

The study featured a 3x10 repeated measures design in which MouseWeight (88, 160 and 232 grams) was a between-subjects variable and Presented distance was a within-subjects variable. The horizontally presented distances were: 1.95, 4.495, 6.95, 9.4, 11.8, 14.3, 16.8, 19.2, 21.7 and 24.2 cm. It should be noted that, due to a technical error, these distances slightly diverged from van Beurden's (2013) presented distances. However, with a maximum difference of 1 mm it is expected that these small differences do not impair the comparability between the results. The vertically presented distances were: 2, 4.5, 5.7, 7, 8.2, 9.5, 10.7, 11.9, 15.5 and 16.9 cm and were unaffected by the error. The dependent variable, estimated distance, was also the same as in the original study. Participants estimated measurements were first averaged and then converted to a computed percentage of the under- or overestimation following procedures based on Waller (1999).

$$Estimation\ (\%) = \frac{Average\ Distance\ Estimation}{Presented\ Distance} \times 100$$

The second set of measurements was gathered in a post-experimental questionnaire. This questionnaire consisted of two separate parts. The first part followed a funneled procedure to assess participant suspicion of the experimental goals and manipulations. The questions consisted of a mix of open questions, seven point Likert scale questions and a combination of both. In the last case the open component was to allow participants to further elucidate their Likert scale answer. The second part of the questionnaire was derived from the AttrakDiff2 user experience questionnaire (Hassenzahl, 2007a). The questionnaire is based on a model that distinguishes between pragmatic and hedonic product quality. The amount of quality perceived in each is contingent on the extent to which it supports 'do'-goals and 'be-goals' respectively (Hassenzahl, 2007b). The original AttrakDiff2 questionnaire contains 28 items across four dimensions: Pragmatic Quality (PQ), Hedonic Quality-Identification (HQI), Hedonic Quality-Stimulation (HQS) and general attraction (ATT). The questionnaire items from the original were evaluated within the scope of mouse apparatus evaluation and because some items were deemed inappropriate in this context, the questionnaire was adapted to contain only 13 items. Moreover, since the participants all spoke Dutch, the word-pairs were translated in Dutch using both the English and German versions as a reference. The full questionnaire can be consulted in Appendix B.

Scale	English	Dutch
HQI	tacky - stylish	stijlloos – stijlvol
	cheap - premium	goedkoop - waardevol
HQS	conventional - inventive	conventioneel - origineel
	conservative -innovative	conservatief - innovatief
	ordinary - novel	gewoon – ongebruikelijk
PQ	complicated - simple	ingewikkeld - simpel
	impractical - practical	onpraktisch - praktisch
	unpredictable - predictable	onvoorspelbaar - voorspelbaar
ATT	disagreeable - likeable	onaangenaam – aangenaam
	ugly - attractive	lelijk – mooi
	bad - good	slecht - goed
	repelling - appealing	afstotend - aantrekkelijk
	discouraging - motivating	ontmoedigend – motiverend

Table 3 Translated word pairs used and their English counterparts (full translation sheet in Appendix A)

Due to the fairly large modifications of the AttrakDiff2 questionnaire, the internal consistency of the scales was first considered (HQI: $\alpha = .614$; HQS: $\alpha = .540$; PQ: $\alpha = .672$; ATT: $\alpha = .847$). Because except for the ATT dimension all dimensions failed to meet the often used .8 criterion suggested for widely used scales (Carmines & Zeller, 1979), a more exploratory approach will be taken and the items selected will be considered both on a per-item basis and in relation to items in the same scale.

3.1.2 Participants

Participants were recruited using the participant database of the J.F. Schouten School for User-System Interaction Research. A total of 90 participants were recruited of which 35 were female. The mean age of the participants was 34.34 *(SD = 19.02; Range = 17-69)*. The experiment took approximately 30 minutes and the participants received an incentive of 5 Euros for their participation.

3.1.3 Task

In the experimental task, participants were to estimate distances between two points presented either horizontally or vertically on the screen. The points were provided by using button pairs that would appear at random offsets on the screen. Each pair consisted of a button labeled 'A' and a button labeled 'B'. The buttons were the same as in van Beurden (2013) and featured a rectangular shape with a width and height that depended on the direction of presentation. In the horizontal presentations the buttons featured a width of 0.5 cm while in the vertical direction the buttons featured a height of 0.5 cm.

The distance estimation process was purposefully complicated because the A and B button would not appear simultaneously on the screen. Thus, when a new distance would be presented only the A button would appear. Participants then had to press this button and move the mouse 1.33 cm in the presentation direction (either horizontal or vertical) to reveal the corresponding B button. After which this button should be pressed and an answer could be inputted in a textbox that would appear (see Figure 1 for the complete sequence).

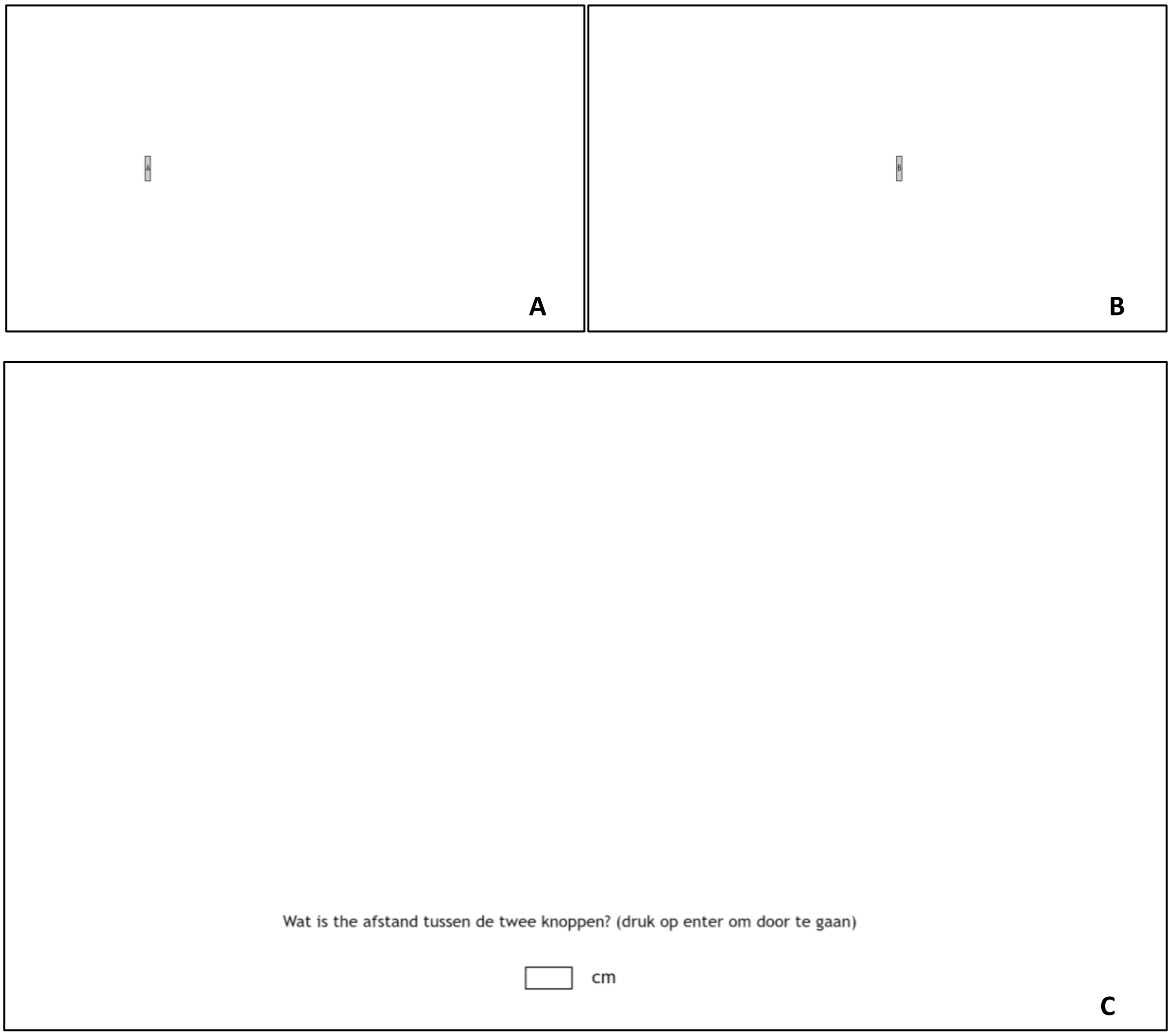


Figure 1 Steps in the distance estimation task order: A: The first button to be pressed, B: The second button to be pressed, C: Filling out the perceived distance between the two buttons.

3.1.4 Setting

The experiment took place at Eindhoven University of technology's IPO building in the PSYLAB. Within the lab three small rooms were used for the participants to partake in the experiment. Two rooms contained a Dell s2209w 21.5 inch, 1920 by 1080 pixels resolution screen and one contained a Dell U2212HMc screen with the same resolution and screen size. Each room, depending on the weight condition, also contained a different version of the apparatus, a computer mouse. During the experiment the experimenter occasionally walked around to observe the participants, answer possible questions, and to be available whenever the participants were instructed to call the experimenter.

3.1.5 Apparatus

In light of the controversy regarding previous work on Effort Theory, the choice for the apparatus, a computer mouse, and manipulations with this apparatus were of fundamental importance. Establishing the specifications of the apparatus for the first experiment required a small pre-study and the pilot to gain insight into the variety of specifications of computer mice and to which extent changing these specifications would be effective. Based on this pre-study, a standard issue computer mouse (Logitech M90) was selected because it could easily be opened and closed and therefore could be fitted with additional weight. Moreover, it had a standard look and feel that matched other mice commonly used in the laboratory and it was fairly cheap.

Since a total of three mice would be used, two of which would have to be equipped with the added weight. The range of weights was determined by a small study on ecologically valid mouse weights. The first modified mouse would be heavier than the heaviest wired mouse found in the ecological validity mouse study and the second mouse was unnaturally heavy and was expected to be salient for the participants and therefore to possibly raise suspicion. The mice weights were as follows: 88 grams 160 grams and 232 grams (including the mouse wire) making the step size 72 grams. Modification of each mouse was done unobtrusively by opening it and adding weight in the form of plastic bags containing lead used in recreational fishing. Foam was added to avoid the weights from moving which could lead to mouse malfunction or auditory or vibration feedback which could inform the participants that the mouse was modified.

Since no internal hardware was changed and standard mouse drivers were available, no specific software was required. The experimental application was developed in Adobe AIR. It was attempted to recreate the original application as closely as possible in terms of functionality. The internal gain was lowered to three across all conditions and the default Microsoft Windows “enhance pointer precision” was disabled such that the gain remained fixed. The lower gain was chosen in order increase the arm movement that participants had to make to move to the on-screen target and thereby have participants experience the weight to a larger extent.

3.1.6 Procedure

Participants that were invited were told that they would participate in an experiment concerning a computer mediated distance estimation task. After participants had arrived, they were welcomed, appointed to one of the three experimental rooms and were asked to sit behind the computer. After reading and signing an informed consent form, both the form and the pen that was used were collected by the researcher to prevent participants from using either as a potential measurement cue or tool. Participants were then instructed to follow the onscreen instructions. Both in the on-screen instructions and in a short briefing by the experimenter, participants were told that when they had questions they could call the experimenter.

Participants were first shown the instructions for the horizontal length estimation task. These instructions were largely based on those used by van Beurden (2013) in the original experiment. However, based on results and observations from a pilot study it was opted to include instructions which emphasized that participants should use only visual input and should follow their first feeling when estimating the distances (see Figure 2). These additional instructions were included because participants in the pilot study tended to use response strategies such as looking at the screen frame and using limbs as measurement aids.

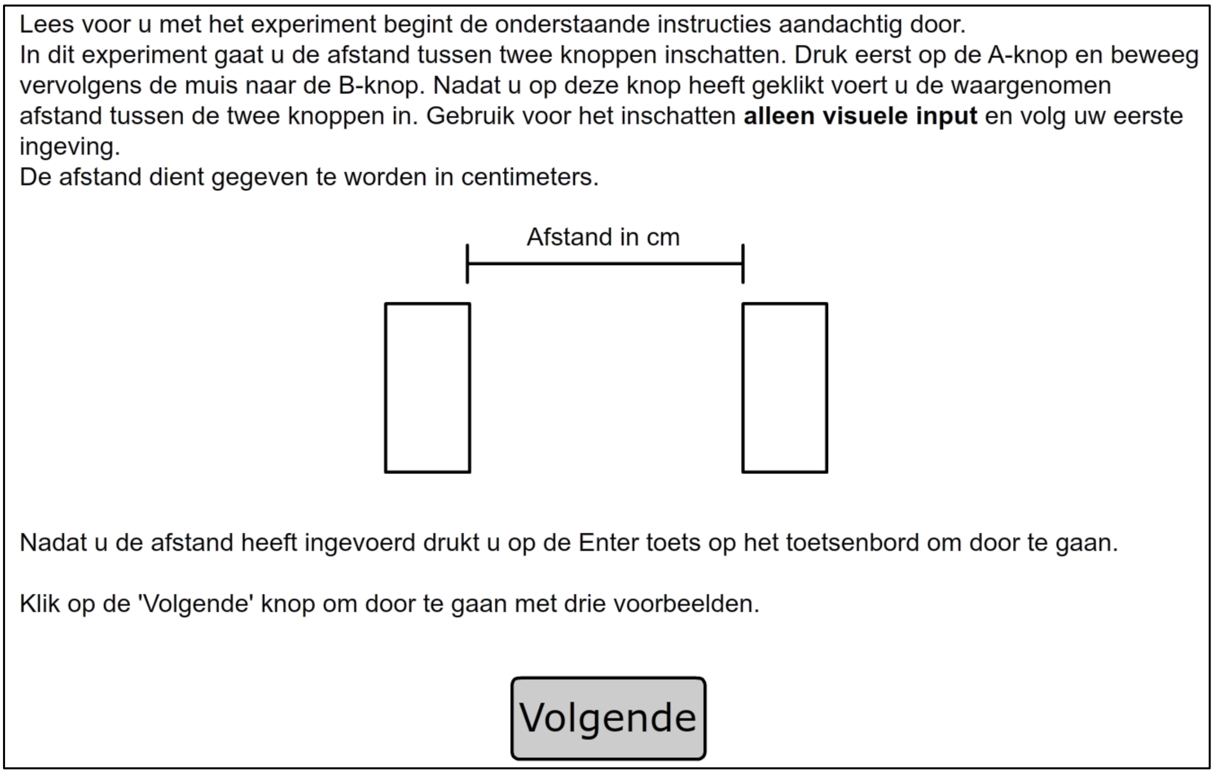


Figure 2 Screenshot of the instructions including the emphasis on the use of visual input and first feeling

After the instructions were read, three practice trials would be performed in which the distance between the buttons was a randomized number within the range of the experimental trial distances. After these trials were completed participants were told that they would continue with the actual experiment. Participants then estimated the length of ten fixed number distances that would be shuffled and repeated three times. In the vertical length estimation task that followed, instructions were largely the same and again, three practice trials were presented before the randomized and repeated trials would start.

When the distance estimation task was completed, the participants were asked to call the researcher. The researcher then came into the room and registered the computer and participant numbers on a questionnaire form. The participant was then escorted out to a large adjacent room and each participant was placed behind a separate table and was asked to fill out the questionnaire. Both within a short verbal explanation, and within the questionnaire instructions, participants were clearly and explicitly requested not to go back to any previous page or question. To ensure compliance, the questions were spread out across multiple, single side printed pages, and the experimenter occasionally observed the participant while taking the questionnaire. When the questionnaire was completed, the participant received the incentive and was taken to a hallway adjacent to the questionnaire room where a debriefing would be done. After being thanked for their participation they were then escorted out.

3.2 Results

Before the distance estimation analysis was done, the data was screened for potential outliers. One participant was excluded due to the gain settings not being properly set. Next, the data was checked for outliers on a per condition level. Two participants were found to have provided extreme estimates across nearly all trials. These participants were deemed distance unaware, or lacking a proper idea of distance, and were therefore discarded. The remaining participants were analyzed by employing the Median Absolute Deviation method (Leys, Ley, Klein, Bernard & Licata, 2013). A conservative rejection criterion of 3 was maintained to determine other outliers. When an extreme outlier was found, the three raw distances that were provided by the corresponding participant were checked for typing errors. When there was one clear extreme in the three provided answers this value was removed (0.06% of all reports) and a new average and percentage estimation was computed using the remaining estimations. In cases where more values were found to be extreme or where there was no clear typing error, the outlier (1.60% of all reports) was recoded into the nearest extreme value (either minimum or maximum) computed using the Median Absolute Deviation method. It was found that some participants were often represented in the outlier calculations. As was done for the less represented participants, the values for these participants were recoded. It was opted to leave these participants in because their percentage estimations were not extreme to the extent that they could justifiably be characterized as distance unaware. In total 4.66% of all cases were modified due to outliers.

One of findings in the original work was that the provided answers were generally overestimations. Therefore, it was first considered whether there was a similar tendency towards overestimation in the current experiment. In a one-sample *t*-test the dependent variable, the distance estimation converted to a percentage, seemed to indicate that this was the case for both the horizontal (*M = 110.73, SD = 27.11; t(86) = 3.69, p = <.001, d = .40*) and the Vertical axis (*M = 122.62 , SD = 30.09; t(86) = 7.01, p = <.001, d = .75*) whereby there was a larger effect size for the vertical presentation overestimations.

The main question in the first experiment was whether the between-subject weight differences influence the distance perception. Like in the original study, the horizontal and vertical distance estimates were analyzed separately. For both the horizontal and vertical distances, a Mixed model ANOVA with Estimates as within-subject and weight as between-subject variable was employed. Weight consisted of the 88 (*N = 29*), 160 (*N = 29*) and 232 gram (*N = 29*) conditions. Based on Mauchly's test it was found that the assumption of Sphericity was violated for both the horizontal and vertical distance estimations therefore, the Greenhouse-Geisser correction was applied in both cases. Note that in addition to often reported effect size measures eta squared (η^2) and partial eta squared (η_P^2), effect sizes for the main analysis were also reported in terms of the generalized eta squared (η_G^2) when differences are found with η^2 or η_P^2, as has been recommended for repeated measures ANOVA (Bakeman, 2005).

The horizontal presentation analysis showed a significant main effect for Estimates, which means that estimation accuracy differed across the presented lengths (*F(2.34, 196.39) = 48.43, p = .000,* η^2 *= .143,* η_P^2 *= .505*). A more detailed look into this effect using post-hoc tests revealed that the shortest distances tended to be overestimated more than the longer distances. The between-subject weight variable did not show a significant effect, nor was there a trend visible that signaled a possible effect of weight (*F(2, 84) = 0.03, p = .971,* η^2 *= .000,* η_P^2 *= .001*). Likewise, the interaction between weight and Estimates was not significant either (*F(4.68, 196.39) = 0.44, p = .805,* η^2 *= .000,* η_G^2 *= .003,* η_P^2 *= .010*) which shows that the effect of weight was not different across different presented distances.

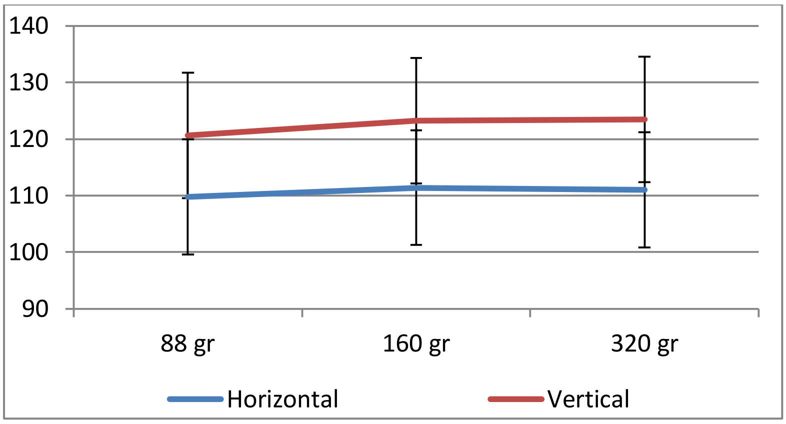


Figure 3 The Main effects of Weight for Horizontal and Vertical estimations (with a 95% CI)

In most ways the results for the vertically presented distances were similar to that of the horizontal distances. A significant main effect of the presented distances was found ($F(3.13, 263.06) = 30.15, p = .000, \eta^2 = .088, \eta_P^2 = .264$) and the lower distances were overestimated more than the higher distances. Both Weight ($F(2, 84) = 0.082, p = .921, \eta^2 = .001, \eta_P^2 = .002$) and the interaction between estimates and weight ($F(6.26, 263.06) = 0.98, p = .444, \eta^2 = .006, \eta_P^2 = .023$) were not significant. Overall, as can be seen from Figure 3, no trend for weight was found.

The second part of the experiment consisted of the post-experimental survey. An important aim of this survey was to determine whether participants were suspicious of the experimental intentions. As the bias assessment part of the questionnaire only concerned participants whose data had been incorporated, the participants that were excluded in the previous analysis remained excluded.

Analysis of the open-ended survey questions was done per weight on a per question basis. Generally, in the first question, a great variety in terms of the extent to which an answer was specific was found. The answers were coded according to the main hypothesis that the participant expressed. When answers did not contain a specific hypothesis, or did not provide a hypothesis at all (e.g. answers such as: “distance estimations”) they were coded as ‘generic’. Some participants discussed multiple hypotheses, in those cases the first specific hypothesis was coded and used, except when one of the hypotheses involved the mouse weight. In those cases the mouse weight hypothesis was always coded. Hypotheses that were specific but were not often mentioned across all conditions were coded as ‘other’. For an overview of all codes used, Appendix C can be consulted.

Across all conditions, when participants provided an experimental goal, it tended to involve a suspected difference between horizontal and vertical distance estimation or an influence of the button position on the screen (i.e. whether buttons were presented more on the left or on the right side of the screen). As expected, within the first condition no participants indicated any suspicions about the involvement of mouse weight (mouseweight) in the experiment. In the second condition, 4% (one participant) and in the third condition 7% (two participants) immediately reported thinking that mouse weight was involved (see Figure 4).

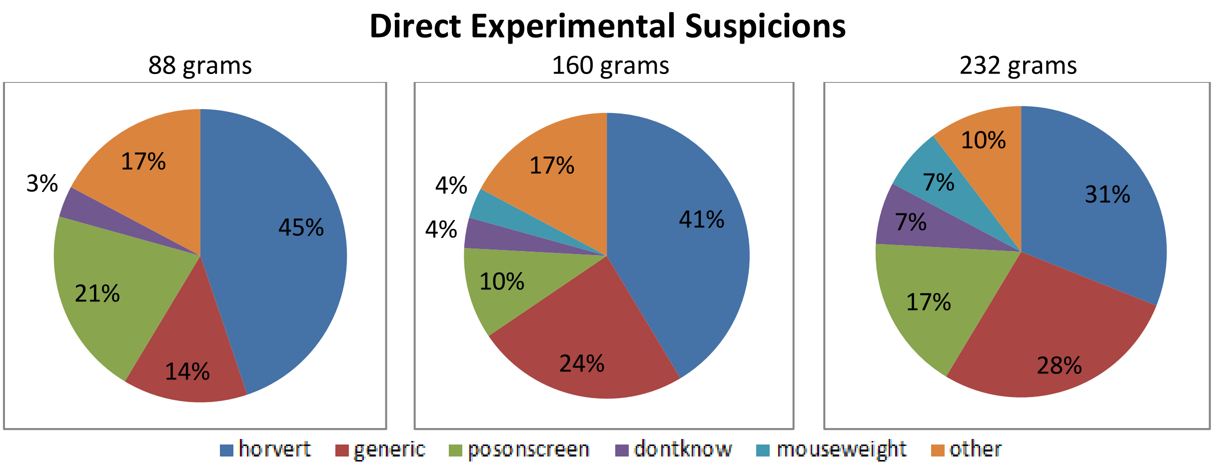


Figure 4 Overview of participants' experimental suspicions per condition immediately after the distance estimation task

The next question within the funneled debriefing was focused on determining how participants experienced their interaction with the mouse. No significant differences between the conditions or an overall trend between the weight conditions was found ($F(2, 84) = 0.97$, $p = .385$, $\eta^2 = .022$). From the strongly varied answers in the open part of this question it could be derived that the measurement may have been noisy. The main cause of the different interpretations seems to be a different understanding of the "mouse" concept. Specifically, some seemed to think of the mouse as a synonym for the on-screen cursor while the question was originally meant to measure the interaction with the physical mouse. The answers to both parts of the question were coded such that any answer above 4 would be interaction as expected, any answer under 4 would be coded as interaction unexpected. When multiple reasons were mentioned, the first mentioned was coded.

It was found that most participants who indicated abnormal mouse interaction across conditions discussed either the gain of the mouse or the fact that the cursor became invisible after pressing the A button. Only in the last condition (232 grams) was the mouse weight mentioned as being out of the ordinary. Some participants mentioned the mouse weight in the explanation but rated the interaction as being normal nonetheless. Figure 5 shows an overview of the interaction ratings.

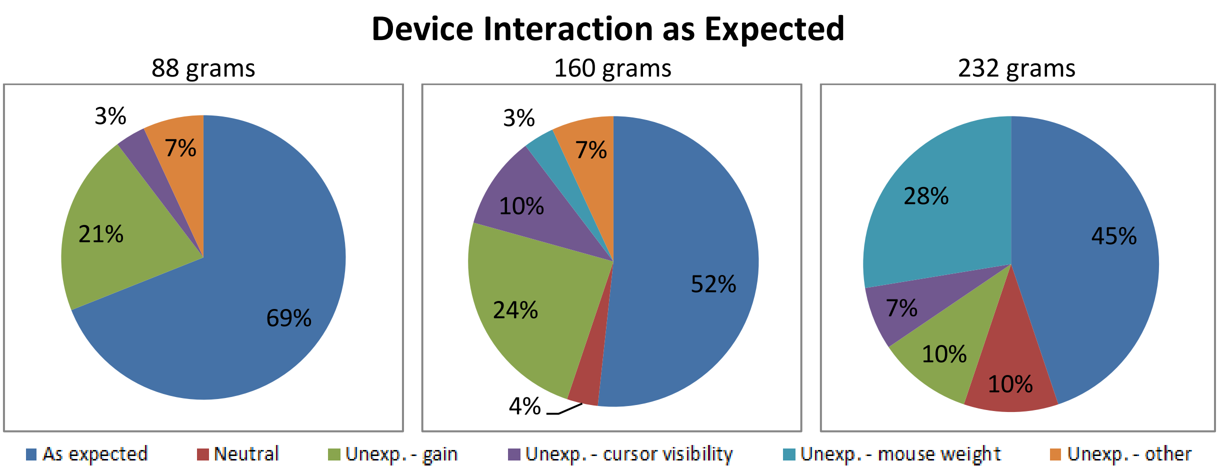


Figure 5 Overview of participants' view on Device interaction per condition

Next, participants were to answer two Likert scale questions that specifically asked participants to rate mouse weight and mouse gain. Participants who received a heavier mouse rated the weight as significantly higher *($F(2, 84) = 18.09$, $p = .000$, $\eta^2 = .301$)*. Post-hoc tests revealed that only the heaviest mouse condition was significantly different from the other two (see Figure 6). No significant differences was found between the first (88 grams) and the second (160 grams) conditions. The gain scores in the scale question were closely considered and no significant gain differences were found between the conditions *($F(2, 83) = 0.29$, $p = .750$, $\eta^2 = .007$)*.

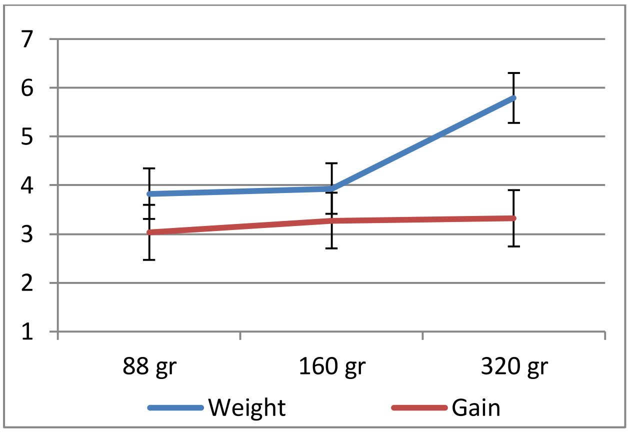


Figure 6 Ratings for perceived magnitude of the Gain and Weight (with 95% CI)

The final question from the funneled debriefing asked whether participants thought their length estimates were impacted by their interaction with the mouse. In all conditions, those who believed that their interactions were impaired, mainly thought that this was the case due to the gain settings (see Figure 7).

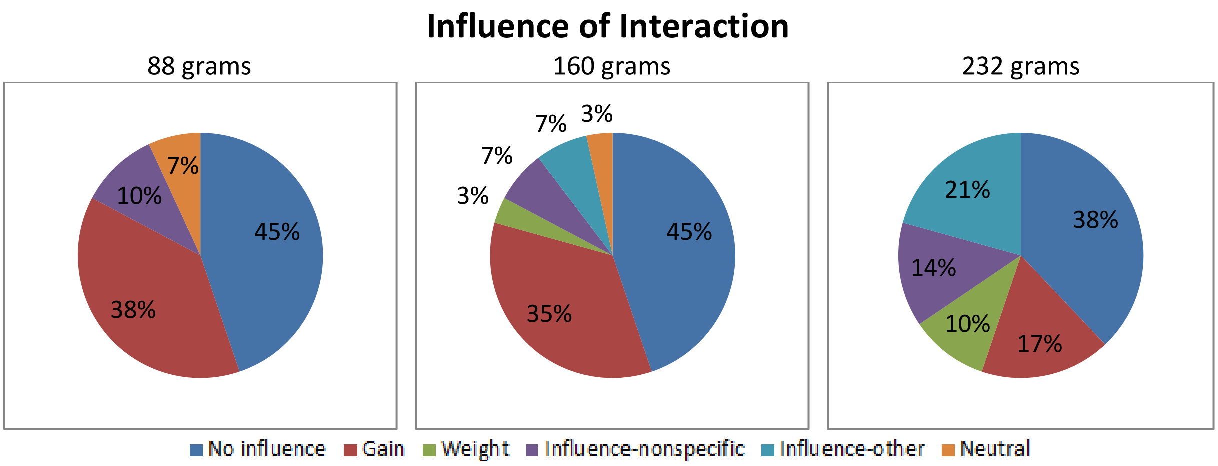


Figure 7 Overview of subjective rating for influence of interaction on estimations per condition

The second part of the questionnaire was aimed at determining how participants in each condition perceived the mouse in terms of hedonic and pragmatic quality, and general attractiveness. In the analysis of this data, only the participant who participated under the wrong gain settings was removed from further analysis.

As noted in section 3.1.1, the reliability scores for the dimensions were too low to extract factors from the items except the score for attractiveness (ATT). For the other dimensions the items were analyzed separately. Each separate item and the ATT dimension were used as a dependent variable in an ANOVA that considered whether the scores for these variables differed per weight condition. No significant differences were found for Attractiveness ($F(2, 86) = 0.79$, $p = .925$, $\eta^2 = .002$) and no trends were visible (see Figure 8). The items within Pragmatic Quality (PQ) showed slight downward trends with weight increases but none was significant (see Figure 8) 1: complicated - simple ($F(2, 86) = 2.74$, $p = .070$, $\eta^2 = .060$), 2: impractical - practical ($F(2, 86) = 2.25$, $p = .112$, $\eta^2 = .050$) and 3: unpredictable - predictable ($F(2, 86) = 1.85$, $p = .163$, $\eta^2 = .041$).

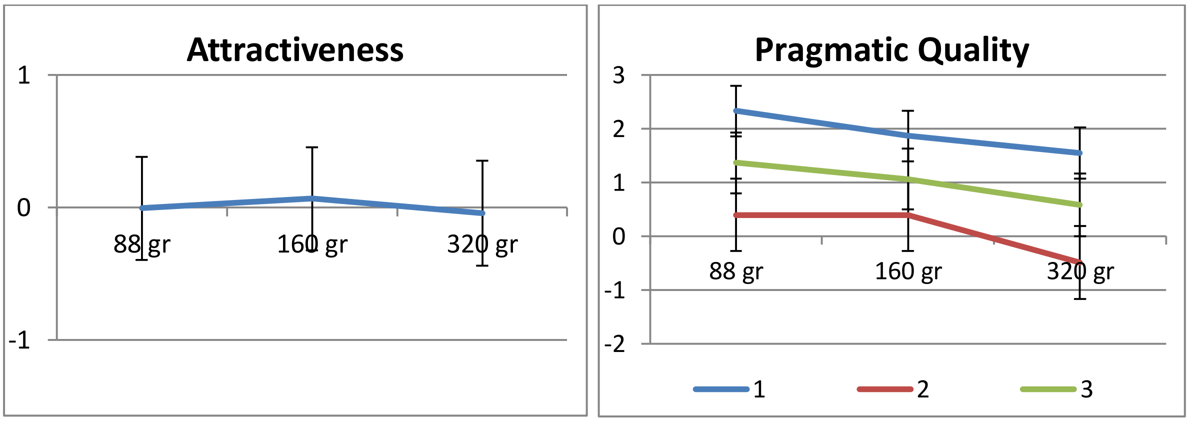


Figure 8 User experience dimension & item scores (with 95% CI). Left: Attractiveness Dimension; Right: Pragmatic Quality items (1. complicated-simple, 2. impractical-practical & 3 unpredictable-predictable) per condition

The Hedonic Quality-Identification (HQI) item cheap-premium ($F(2, 86) = 3.15$, $p = .048$, $\eta^2 = .068$) significantly differed between the conditions while tacky-stylish did not show a clear trend and was not significant ($F(2, 86) = 0.91$, $p = .405$, $\eta^2 = .021$). Hedonic Quality-Stimulation items conventional-inventive ($F(2, 86) = 2.34$, $p = .102$, $\eta^2 = .052$), conservative-innovative ($F(2, 86) = 2.80$, $p = .066$, $\eta^2 = .061$) & ordinary-novel ($F(2, 86) = 4.89$, $p = .010$, $\eta^2 = .103$) all showed an upward trends with weight increases. Only ordinary-novel showed significant differences between the conditions (see Figure 9).

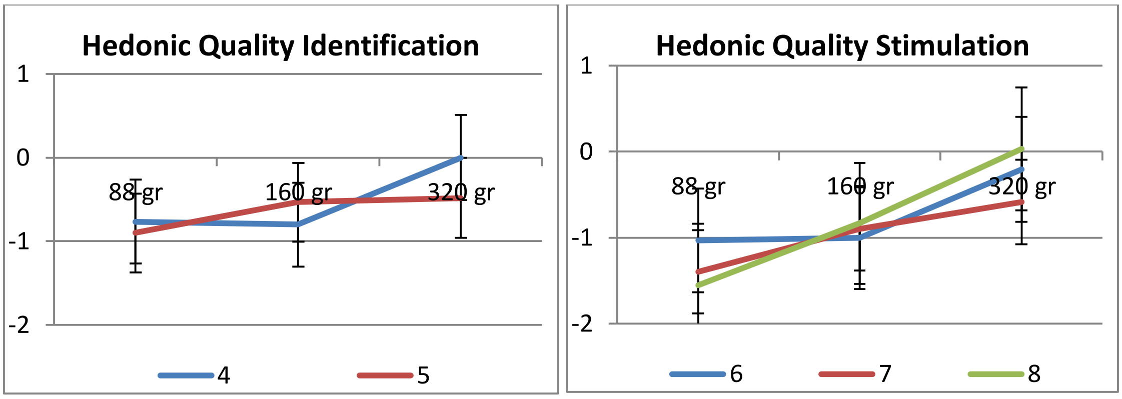


Figure 9 User experience item scores (with 95% CI). Left: Hedonic Quality-Identification (4: cheap-premium & 5: tacky-stylish); Right: Hedonic Quality-Stimulation (6: conventional-inventive, 7: conservative-innovative & 8: ordinary-novel) per condition

From these results it can be seen that, in line with the second hypothesis, there is an indication for lower perceived support of the apparatus in doing the tasks. However, the three HQS items seemed to indicate that participants found the heavier mouse more stimulating. The ATT rating remained the same between the conditions and similarly the items in HQI showed no clear trend for differences in that aspect.

3.3 Discussion

The main goal for the experiment was to consider whether earlier findings regarding the influence of mouse gain on distance estimations (van Beurden, 2013), could be attributed to an effort based explanation. To achieve this goal, van Beurden's original work was largely replicated with an altered manipulation in the form of mouse weight. Additionally, a two-part questionnaire was used to assess potential bias in the responses and to measure participants' user experience of the apparatus. The second part fulfilled an additional experimental goal, determining the effect of apparatus weight on subjective perceptions of the apparatus.

From the results it becomes evident that the current experiment was unable to provide grounds to assume an influence of weight on distance estimations. Therefore, the findings by van Beurden (2013) could not irrefutably be attributed to an effort account based on the current experiment. However, such a mechanism can also not be ruled out. Some potential issues regarding the employed methodology could have influenced the results such that an effort related explanation was not discovered.

One of the first analyses was done to form a general image of the estimates per direction. It was considered whether participants were able to correctly estimate the distances, or whether they generally tended to under- or overestimate. For both vertical and horizontal estimations it was found that a significant overestimation took place. The effect sizes found were slightly smaller than in van Beurden (2013) but followed the same direction, i.e. the vertical estimations were overestimated to a larger extent than were the horizontal. The fact that the effect sizes were smaller is interesting. The current experiment used a smaller screen and therefore smaller frame, previous findings indicate that with smaller frames, higher length estimations tend to be provided (Künnapas, 1955 as discussed in Thompson & Schiffman, 1974), based on this notion the difference is unexpected. Potentially it is a consequence of the manipulation that has been used. Indeed, the gain and weight manipulation are difficult to compare in terms of required effort, it could be that the weight manipulation was less effortful and therefore led to less overestimations. Additionally, the way in which participants estimated the distances could have been of influence. Irrespective of what mechanisms have influenced the result, and despite the small differences in effect size, a general distance overestimation is in line with previous findings (for a discussion see Hamburger & Hansen, 2010).

The effect sizes of weight on distance estimations, both horizontally and vertically, were negligible and no apparent trend could be discovered. There are several possible reasons for this result, some of which concern limitations in the current methodology. Perhaps one of the most likely problems lies embedded in the setup that was used which has proven to be sensitive in several ways. Its sensitivity was first witnessed in an early pre-study by van Beurden (2013), who found that when the two buttons were simultaneously present on the display, participants tended to make estimations before moving the mouse. Therefore, precautions were taken in the later experiment such that visual pre-movement cues were removed from the screen. The B button would appear only after having clicked the A button and the mouse would disappear after the A button was clicked, only to reappear after moving it a certain distance.

In the pilot study of the current experiment, similar pre-cautions were taken as in the original study. However, based on observations and a verbal debriefing with the participants, this approach was found to be insufficient to prevent participants from using response strategies. Several often used strategies were discovered, that had not previously been highlighted. Some participants placed their fingers on the screen or backtracked with the cursor to the A button positions to allow estimations to be made between two simultaneously present points on the screen. Based on these findings the instructions for the first experiment were adapted in an attempt to accommodate for this behavior. The instructions additionally stated: "use **only visual input** and follow your first intuition" (bold as in instructions, instructions translated from Dutch). However, observations and participant debriefings again showed that participants used response strategies. In fact, usage of limbs as improvised rulers was newly introduced as a response strategy. Irrespective of the chosen strategy, it is expected that estimated distance may be lower as these measurements are independent of mouse movement and therefore insensitive to the mouse weight manipulation.

Considering these two easy ways in which the experimental procedures could be ignored despite counter measures, it should be concluded that the current setup is likely to remain highly sensitive to deliberate participant disregard of the instructions. Because some participants were observed using strategies or reported doing so in the debriefing, a list could be made that would allow for a test of the strategy explanation. Using this list it was tested whether weight differences could be discovered after excluding 16 strategy employing participants, which was found not to be the case. However, the strategy explanation should not be ruled out based on this fact. Given that it is impossible to continuously observe three participants simultaneously, it is expected that the list is incomplete.

Another potential reason that no differences were found was because the presented weights were too similar to find a difference. Some of the data from the questionnaire could be interpreted as supportive for this possible reason. Specifically, the finding that even in the highest weight condition, few participants initially discussed the weight as relevant in the experimental goal could mean that too little effort was induced. Moving around a computer mouse is a supported action as both the arm and the mouse itself are supported by the surface that the mouse is positioned on. Moreover, the bottom of the mouse featured small plastic nubs designed to sharply reduce resistance when moving the mouse around. While the gain was reduced to make participants experience the weight more through larger movements, it could be that the weight was not experienced to a sufficient degree to have an effect. The potential of too small weight differences forms a problem for a possible replication and improvements of the current study, as the current apparatus, through its physical dimensions, constrains the amount of weight that can be added. With the weight that was installed in the mouse, the usable empty space inside (empty space with no underlying electronics) the mouse was close to being filled up.

Finally, there is a theoretical possibility that, because few participants expected an effort based manipulation, no effect was found. If previous effects occurred due to demand characteristics, lack of participant suspicion could explain the fact that no trend was found. However, in light of the other two limitations and possible explanations this account seems the least likely and could therefore not reasonably be defended.

Despite that no effect of weight was found that could further be researched using the post-experimental questionnaire, the questionnaire remained interesting because it could provide a perspective on why the manipulation did not achieve the expected effect. Directly after the distance estimation task, all participants tended to think that the experiment involved finding differences between horizontal and vertical distance estimations. Even in the highest weight condition, only a very small percentage (7%) immediately suspected that mouse weight was involved. As noted in the weight manipulation limitation discussion, this finding could indicate that the weight manipulation was too small. However, while it may have been too small to achieve an effect, based on more focused questions it seems that the weight was at least noticed by participants in the heaviest weight group. A fairly large percentage (28%) in that group indicated the weight as being unexpected (Q2) and significantly higher than other groups (Q3). However, only a fairly small percentage (10%) subjectively felt that weight had impacted their estimations. The high weight rating in the third condition in relation to the low subjective influence of influence of weight seems like an interesting paradox. There are several possible reasons that could explain this finding. One, again, lies in response strategies. When such strategies were applied they could be expected not to strongly influence distance estimates but while moving the mouse the weight of the mouse could still be experienced. Another possible explanation is that the mouse weight, during normal interaction, was not experienced as heavy at all but, due to the low gain, for longer distances participants may have lifted the mouse and experienced the weight during the act of raising the mouse. Some participants in the debriefing verbally reported that with longer distances, they occasionally lifted the mouse.

Based on pilot findings which indicated that the perception of the mouse may be different across several conditions, the second part of the questionnaire assessed the perceived mouse quality in terms of hedonic and pragmatic quality. Due to the low internal consistencies for three of the scale, the items in these were considered separately. Items pertaining to PQ and HQS were found to show non-significant trends with weight increases. The PQ items showed a decrease for increases of weigh which was in line with the hypothesis 2. However, none of the trends for the PQ items were significant. For HQS the direction of the trend was upwards with weight increases. One of the items showed a significant difference between the weight groups, ordinary-novel was significantly higher when the apparatus was heavier. The findings for HQS indicate that the mouse may have been perceived as more stimulating. A possible reason could be that with the added weight the mouse was seen as special, perhaps even as meant for use by a special group, similar to what a trend indicated may have also been the case in the pilot study. As for the HQI items, the cheap-premium item showed significant differences between the weight groups such that with weight the apparatus was seen as more valuable. However, since the other HQI item (tacky-stylish) did not show any differences or trends, no HQI differences should be inferred. Finally, the ATT dimension showed no differences or trends meaning that the attractiveness of the mouse did not differ per condition.

While these findings are interesting, there are some limitations that should be brought in the limelight. First the choice of removing items deemed inappropriate within the context of rating a mouse apparatus. The HQS, HQI and PQ could not be extracted from their respective selected items for two reasons. First, the internal consistencies of the items that made up the dimensions were low and while they showed the trends in the expected directions, the trends were generally weak. Second, on a more conceptual level it is unclear how the removed items change the meaning of the dimension. Both prevent that any solid statement can be made about the dimensions and regarding hypothesis 2 and 3. The analysis for these items should therefore strictly be considered as indicative for how their dimensions could be affected by effort. Despite a seeming lack of significant differences for the weight groups on the AttrakDiff2 dimension items, the results are not uninformative. It could be another indication that the manipulation was not adequate for the goal of measuring effort differences. A second limitation involves the context set by the preceding parts in the questionnaire. In the first part of the questionnaire participants were asked about several specific mouse characteristics (such as weight and gain) which could have biased participants' memory of the mouse or otherwise have made some characteristics more salient during the rating. This in turn could have had an influence on the mouse appreciation. A final limitation is that the current findings are expected to be highly specific for the current setup and settings. In the debriefing several participants indicated that the mouse was sluggish, a judgment which was likely due to the gain setting. Because the mouse gain and the 'Enhance pointer precision' setting were changed to non-default states. It is unknown whether participants included an appreciation of the gain setting in the mouse appreciation or whether they saw gain as a setting and did not incorporate it. Due to this unknown factor, the current findings should arguably not be generalized beyond the current setup and settings.

4. Experiment 2 – Effect of Controller Weight on Distance Perception

The results from the previous experiment were not in line with the effort hypothesis as no effect of apparatus weight on exocentric distance estimations could be discovered. An evaluation of the post-experimental questionnaire and observations during the experiment seemed to indicate a possible influence of inherent disadvantages of the setup and apparatus on the results. MEs in the form as used in the previous experiment appear to be too fragile in terms of response strategy usage. Moreover, employing a computer mouse as a manipulation, restricted the effort manipulation for two reasons. First, the mouse's potential for adding weight was too limited. Second, the device is supported on a surface which may further reduce the required effort of using it.

In light of these limitations, within the current experiment it was decided to use a VE and extend Experiment 9 by van Beurden (2013). Not only would adopting the methodology again allow for comparisons with the earlier results, but more importantly, the setup would address most concerns related to the previous experiment. Indeed, employing a head mounted display prevents strategic usage of a screen frame and finger placement on the display. Moreover, the flexibility of the motion tracking system allows for greater apparatus flexibility. Not only can trackers be built in any form and shape, but these devices may also be less likely in evoking suspicion due to the weight as participants are expected to have less experience with this category of devices.

One other important alteration was made with respect to the previous experiment. Due to the poor comparability of the previous user experience measurements and the different way in which the apparatus would be presented in the current experiment, the user experience measurement approach was altered. While in the first experiment the user experience of the apparatus itself was studied, the current experiment will shift the focus towards the evaluation of the interaction with the apparatus. While one could argue that this focus shift should lead to a reconsideration of the user experience hypotheses, past literature suggests that this may not necessarily be the case. Factors concerning system appearance and form have been identified as potentially relevant for broader user experience (Mahlke, 2007). More specifically, Jordan (2000) suggested that physio-pleasure, or the feel of a product in the hand, during interaction may be relevant for products such as telephone handsets and remote controls. Based on these beliefs it is expected that experience of a device may carry over to experience of the interaction itself.

Like in the first experiment, in this experiment it will again be researched whether a device based effort induction influences visual distance perception and general measures of user experience.

4.1 Method

4.1.1 Design and Measures

The study featured a mixed model design in which tracker weight (132 and 542 grams) was manipulated between-subjects and the distances were manipulated within-subjects. Distances were presented in both an exocentric and an egocentric manner. Exocentric distances consisted of representations both on the x-axis and the z-axis. For each axis ten different distances (5, 10, 15, 20, 25, 30, 35, 40, 45 & 50 cm) were presented, the exocentric distance variable thus consisted of twenty distances. Distance presentations each featured randomized offsets on the presentation axis i.e. the distances could randomly be more to the left/right or front/back from the participants' perspective. Egocentric distances consisted of five distances (40, 50, 60, 70 & 80 cm). Each of the distances was presented twice leading to two verbal reports per distance. These reports were registered after which an average was computed. Using the formula previously described in section 3.1.1, this average was converted to a percentage of under- or overestimation which functioned as the dependent variable.

The second set of measurements again consisted of a post-experimental questionnaire series. The first part consisted of the same funneled questionnaire used in Experiment 1. The second part consisted of a single physical fatigue question based on the Borg CR-10 scale (Borg, 1990) which was included as a manipulation check. Finally, the third questionnaire was the AttrakDiff2 user experience questionnaire (Hassenzahl, 2007a). One of the problems within the previous experiment was that some word pairs seemed inappropriate to describe user experience of a device. Because in the current experiment the user experience focus was changed, this issue was no longer present. Therefore, the full Dutch translation of the Attrakdiff2 questionnaire was used in the current experiment (see Appendix A).

The Dutch translation of the Attrakdiff2 scale reached acceptable levels of internal consistency for three of the four dimensions (Cronbach's alpha's: HQI = .801; HQS = .840; ATT = .835[5]. While the PQ dimension (α *= .786*) scored slightly below the .8 threshold, it was kept intact to maintain comparability with other experiments.

4.1.2 Participants

Participants were recruited using the participant database of the J.F. Schouten School for User-System Interaction Research and convenience sampling. A total of 43 participants were recruited. However, for various reasons some had to be disqualified. One participant had participated in Experiment 1, eight participants were partially or fully stereo blind and one participant was excluded due to cancelation of the participation.

Out of the remaining 33 participants, 16 were female. The mean age of the participants was 21.45 *(SD = 2.75; Range = 18-29)*. All participants had a stereo acuity of above 40 second of arc. The experiment took approximately 30 minutes and the participants received an incentive of 5 Euros for their participation.

4.1.3 Task

In the experimental task, participants were again to estimate distances between two points. Since the original VE and geometry from van Beurden (2013) were available, these were also used in the current experiment. In the exocentric presentations both points were provided using a Rubik's

[5] Due to an error in the questionnaire the ATT dimension unpleasant - pleasant and disagreeable - likeable items were omitted from the final ATT dimension calculations.

cube (5.7 x 5.7 x 5.7 cm) that would appear on top of a virtual table. Like in Experiment 1, the task was complicated by not presenting both cubes at the same time. When a new distance would be presented, participants used a tracking device (see section 4.1.5 for more on the device) to navigate a small white sphere (0.5 cm diameter) which functioned as a pointer into a red sphere (2.5 cm diameter) after which the first Rubik's cube would appear. Next, participants were to move the pointer sphere into the first Rubik's cube after which the cursor sphere would become green colored and a sound would be presented. Participants then pressed a button on the tracking device after which both the first cube and the pointer sphere would disappear. The second cube would become visible after participants moved the cursor sphere 3 cm away from the location of the cube. After the second cube was selected and clicked on in a similar fashion, participants would be displayed a text which prompted the participant to provide a verbal estimation of the distance between the cubes.

The egocentric distance estimation task was largely the same as the exocentric task however, in this task only one Rubik's cube was used. After the red ball was touched by the pointer sphere, one cube became visible and had to be selected. When the cube was selected participants were to provide the perceived egocentric distance to the cube.

4.1.4 Setting

The experiment took place at Eindhoven University of technology's IPO building in the Virtu/e lab. The setting was adapted to be similar to the setting used in Experiment 9 by van Beurden (2013). Both the lab itself and an adjacent room were used for the experiment. The adjacent room was well lit and consisted of desks placed against the walls and desk chairs. On one of the tables the informed consent and stereo acuity test were placed. The lab itself consisted of one large room which, in the center, contained a chair placed behind a table on which the participant was to sit down. On one end of the table an instruction form was placed and on the other head a small laptop was placed to which a wired microphone was connected. The room also contained a NVIS nVisor SX111 Head Mounted Display (HMD) with 102H x 64V degrees FOV (111 degrees diagonal), and a resolution of 1280x1024. For each eye a different view was presented that had an overlap of 50 degrees (66%). Both differently weighted tracking devices were also present. However, these devices were covered so that they were not visible to the participants upon entering the lab.

Figure 10 Overview of the Lab environment and demonstration of the apparatus usage

4.1.5 Apparatus

Because the device would form the manipulation there were two important criteria it had to adhere to. First, it had to be comfortable to hold for the participant, and second, it had to have sufficient available room for a weight manipulation. To accommodate for both criteria the device was tailored for the current experiment (see Figure 11). The base of the tracker device consisted of a 20 cm (height) by 4 cm (diameter) PVC tube. Weight in the form of recreational fishing lead wrapped in foam was placed in the lower 14.6 cm of the tube, the part which was to be held by participants. The foam prevented the weight both from moving and making noises when the tracking device was used. The top and bottom of the tube were enclosed using caps. At 5.6 cm from the top a button was added while on the top enclosure cap an LED marker was added. Wiring for both the LED and the button functionality ran through the base to the bottom of the apparatus towards the Phidget interface kit (button) and an LED driver (Tracking LED).

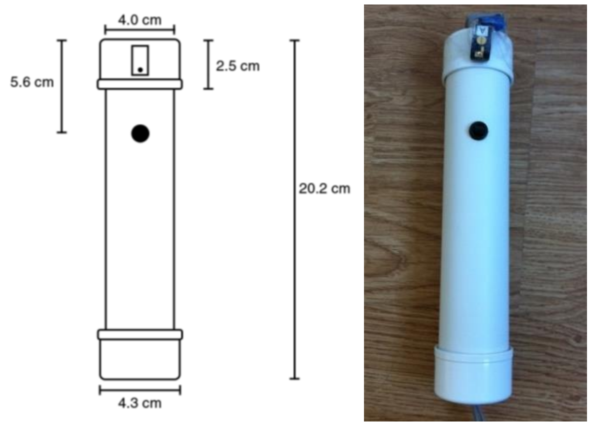


Figure 11 Left: Schematic of the tracker devices; Right: Picture of one of the two tracker devices

The LED was tracked by the camera system in the lab and upon displacement of the LED allowed for pointer actuation. When participants wanted to select an object in the environment they pressed the button using their thumb. After a button press, a signal would be sent to the Phidget interface kit which converted the signal into input for the experimental application.

4.1.6 Procedure

After arrival at the Virtue lab participants were first brought to a room adjacent to the lab. Here they were asked to read and sign an informed consent form. Next they received a short briefing in which they would be told that they would participate in a VE experiment and that whenever they experienced physical discomfort (e.g. nausea or headaches) that they could stop their participation at any time. Finally, before going to the lab room, the Titmus Fly test was conducted to determine participant's stereo acuity.

Upon arrival in the lab room, participants were asked to have a seat on the chair behind the table in the center of the room. On the table a form was placed which contained both a small cover story for the experimental goal, and an explanation for the first task of the experiment (see Appendix D). Like in the first experiment, participants were instructed to "follow [their] first intuition" with the aim of preventing extensive measuring attempts or response strategy usage. In addition, participants were told to move the cursor directly into the second cube to prevent backtracking with the cursor. While participants were reading the form, the environment was prepared for use. After participants finished reading there was an opportunity to ask questions when the task was not understood. Upon having answered possible questions, the form was taken by the experimenter. Then participants were explained that they would be requested to clip a small microphone to their collar such that measurement estimates could be recorded during the task. After participants had given their verbal consent for the audio recordings and had put on the microphone, the recording would be started. Next the HMD would be placed on their heads and adjusted to be stable and comfortable. Participants would then be told that the tracker device would be handed to them and asked to hold out and open their dominant hand. The device was then placed in their hand and verbal help was given to find the button with their thumb on top of the device. After the experimenter took place behind the computer, the first task was initiated.

First the exocentric distance task would be done. For this task either x- or z-axis distances would be presented first, after which the other followed. For all tasks and directions two examples would first be provided. In the first example of the first task, if necessary, the experimenter guided the participant through the trial to ensure understanding of the workings of the controller. The second example trial in nearly all cases was done independently. After the first two example trials were finished the exocentric task commenced. When the exocentric task was complete participants received a brief verbal explanation about the egocentric task after which it was directly started. After both tasks had been completed, first the tracker device would be taken from the participants and covered, then the HMD would be removed from the participants' heads and the audio recording would be stopped. Participants were then asked to take place behind the laptop on the other end of the table on which they filled out the post-experimental questionnaire. When the questionnaire was completed, the participants were debriefed and received the incentive. After being thanked for their participation the participants were escorted out.

4.2 Results

A small number of the verbal reports for which the recorded audio had to be consulted remained unclear or ambiguous. These reports (0.18% of all reports) were therefore considered as missing data and the remaining report was treated as the average response. After these small mutations were done and before the distance estimation task analysis was conducted, the data was first screened for outliers using the Median Absolute Deviation method (Leys et al., 2013). As in the previous experiment, the relatively conservative rejection criterion of 3 was maintained. When an outlier was found, the value was recoded into the nearest extreme value (either minimum or maximum) that followed from the Median Absolute Deviation calculations. After the outliers (2.79% of all cases) were adapted in this way, subsequent normality analyses per condition and presented distance were done.

Many studies in the past, including van Beurden (2013), have reported a tendency towards underestimation of distances in VEs. It was therefore first considered whether a similar effect could be found in the results of the current study. A one-sample t-test against the veridical score confirmed an underestimation effect both for the exocentric x-axis ($M = 86.55$, $SD = 22.63$; $t(32) = -3.41$, $p = .002$, $d = -0.60$) and for the egocentric presentations ($M = 68.21$, $SD = 22.59$; $t(32) = -8.08$, $p = <.001$, $d = -1.43$). For the exocentric z-axis presentations no significant difference was found ($M =$

91.07 , SD = 27.03; t(32) = -1.90, p =.067, d = -0.34). Considering the effect sizes, the pattern of these results closely resembles van Beurden's findings (x-axis: 0.76; z-axis: 0.37; egocentric: 0.96) as the z-axis direction also features a smaller effect size compared to the other presentation directions. A notable difference is the fact that for the z-axis exocentric presentation no significant difference was found. However, this may be explained by the fact that the effect is smaller and the sample size in general was relatively small. Since the direction of the difference is according to expectations, it is expected that with a somewhat larger sample a significant difference from veridicality for the z-axis presentations could also be found.

To examine the first hypothesis, a similar analysis approach as in the first experiment was taken. A mixed model ANOVA was conducted using the average distances converted to percentages of under-/over-estimation as the dependent variable, and the between-subject controller manipulation (132 grams, N=17; 542 grams N=16) as the independent variable. Sphericity violations were found for both presentations and Greenhouse-Geisser corrections were therefore applied.

A mixed ANOVA using all exocentric distances as the dependent variable showed a significant difference between the various presented distances (*F(5.37, 166.43) = 4.69, p = <.001,* η^2 *= .062 ,* η^2_G *= .063,* η^2_P *= .131*). Post-hoc analysis using Bonferroni corrections, revealed that only the highest z-axis distances (45 and 50 cm) differed significantly from the lowest z-axis distance (5 cm). No significant main effect for apparatus weight (*F(1, 31) = 0.26, p = .616,* η^2 *= .004,* η^2_G *= .005,* η^2_P *= .008*) was found (Figure 12), nor was there an interaction between the exocentric distances and apparatus weight (*F(5.369, 166.43) = 1.26, p = .282,* η^2 *= .017,* η^2_G *= .018,* η^2_P *= .039*). Except for the z-axis presentations in the heavy condition, which showed a slightly upward trend, no other trends could be discovered (See Figure 13 & Figure **14**).

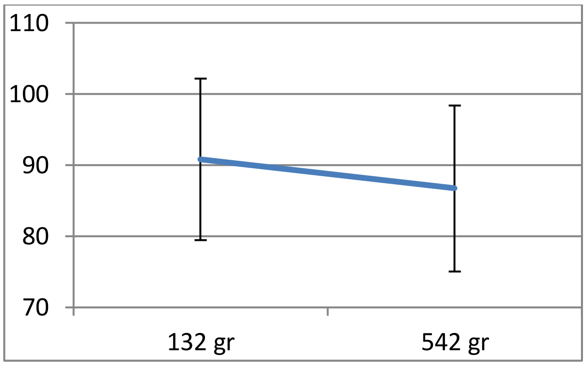


Figure 12 The influence of Weight on distance estimation accuracy for exocentric distances (with 95% CI)

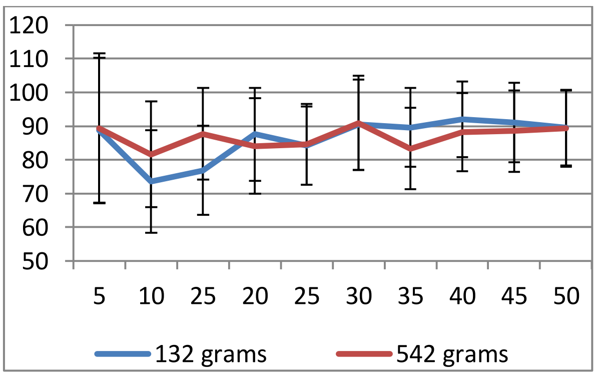


Figure 13 The effect of weight over the presented x-axis distance presentations (with 95% CI)

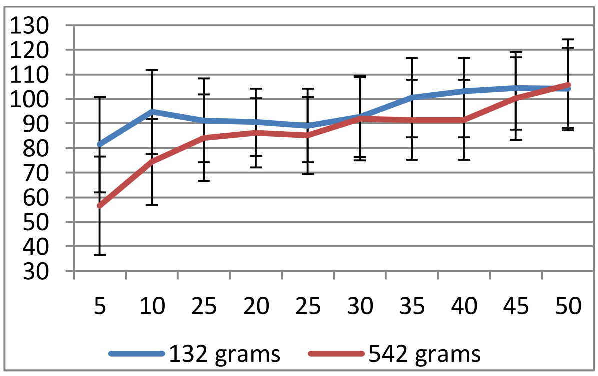


Figure 14 The effect of weight over the presented z-axis distance presentations (with 95% CI)

The results for the egocentric distances were similar to those of the exocentric distances. There were significant differences between the presented distances (*F(2.07, 64.02) = 50.93, p = <.001,* η^2 *= .169,* η^2_G *= .170,* η^2_P *= .622*). Post-hoc analysis using Bonferroni corrections showed that significant differences existed between all distances except for the 40 and 50 cm. Like was the case for the exocentric distances, no significant differences were found for apparatus weight (*F(1, 31) = 0.13, p = .724,* η^2 *= .003,* η^2_G *= .004,* η^2_P *= .004*) and for the interaction between the distance and the weight variable (*F(2.07, 64.02) = 1.29, p = .282,* η^2 *= .004,* η^2_G *= .005,* η^2_P *= .040*). As can be seen in Figure 16, for both weight conditions a trend was visible in which far space was seen as closer to veridicality than near space.

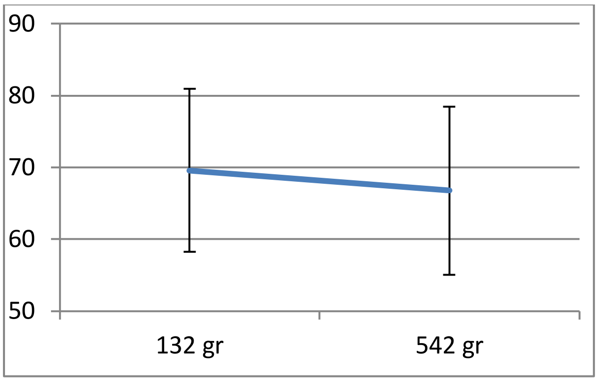


Figure 15 The influence of Weight on distance estimation accuracy for egocentric distances (with 95% CI)

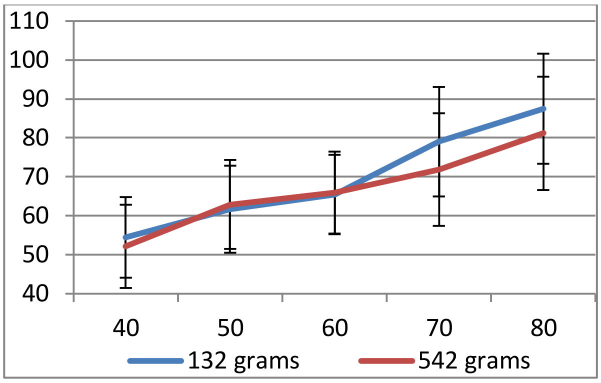


Figure 16 The effect of weight over the presented egocentric distance presentations (with 95% CI)

Based on these findings from both the exocentric and the egocentric presentations, hypothesis 1 cannot be accepted.

The second part of the experiment again consisted of the post-experimental questionnaire. The aim of this questionnaire was to assess possible participant suspicion of the goals, determine participant fatigue and provide insight into general user experience of the interaction with the device. The analysis of the results was done in this order and on a per weight group basis. Answers were coded in a similar way as the previous experiment, for an overview of all codes used Appendix C can be consulted. The participants removed from the distance estimation task analysis remained excluded for all parts of the questionnaire analysis. This was done to prevent possible interaction difficulties experienced during the experiment caused by reduced visual acuity from being attributed to the interaction itself.

As previously noted, participants were provided with an experimental goal in the instruction form that was read before the experiment started. It was expected that this goal would be experienced as convincing and beliefs regarding the intention were thus expected to be derived of those provided in the instructions. Question one considered whether this was indeed the case by asking participants to describe their beliefs involving the experimental aim. Among those who provided a specific goal or hypothesis, often mentioned goals were benchmarking the VE (VEBench) and simply measuring perception in VEs (distInVE). Interestingly, however, even in the first question, in both conditions slightly over 10 percent already believed that embodiment (embodiment) played an important role in the experiment. In both conditions participants believed that the extent to which they were required to move the controller and thus their arm could influence their distance estimates.

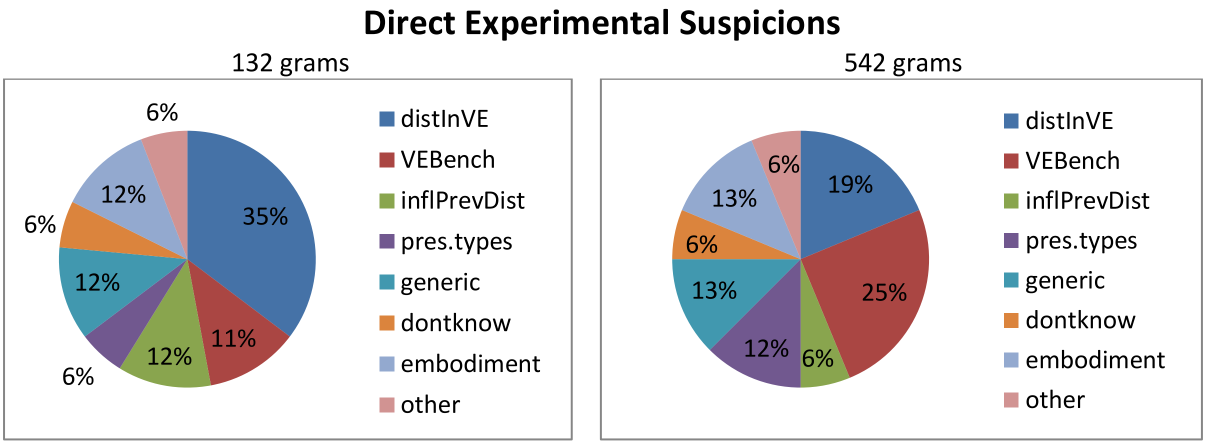


Figure 17 Overview of participants' experimental suspicions per condition immediately after distance estimation task

The second question asked participants to indicate whether the interaction with the motion controller was in line with prior expectations. Based on the Likert item, no differences between the weight conditions was found (*$F(1, 31) = 0.58, p = .453, \eta^2 = .018$*). A consideration of the open part of this question showed that, in both conditions, most participants noted the interaction to be in line with expectations. Some had experience with commercially available motion controllers and the comparison with the Wii remote provided in the instruction form was seen as appropriate. Among those who did report conflicts between prior interaction beliefs and the actual interaction, the inclusion of depth (Unexp. – depth) and the gain setting (Unexp. – gain) were named as unexpected factors. Most interestingly within the context of the experimental goal, however, was to consider whether the controller weight was seen as unexpected. Notably, only one participant in the weighted device condition (6%) explicitly reported the interaction as unexpected due to the controller weight.

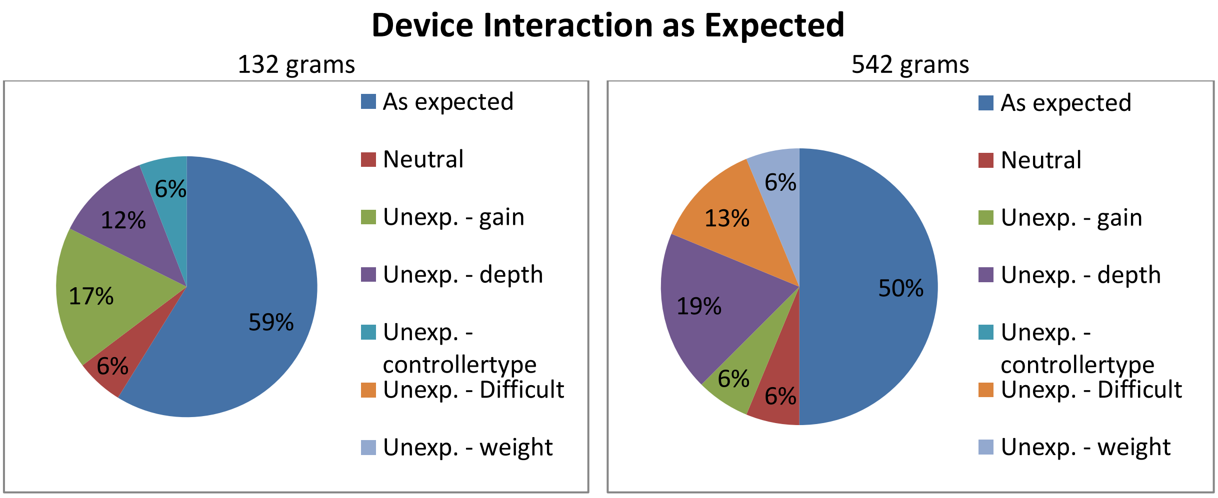


Figure 18 Overview of participants' view on Device interaction per condition

Upon being asked about the apparatus weight and gain, as expected, a significant difference was found for weight (*$F(1, 31) = 10.52, p = .003, \eta^2 = .253$*) while gain did not significantly differ (*$F(1, 31) = 0.04, p = .947, \eta^2 = <.001$*). Finally, it was considered whether participants believed that their experience of interacting with the device affected their measurements. No significant differences were found between the conditions (*F(1, 31) = 2.51, p = .123,* η^2 *= .075*). The additional explanations provided, show that participants in both conditions believed in an influence of embodied cues. In particular, the amount of reaching was reportedly incorporated in estimates. In both conditions, none of the participants referenced the apparatus weight as a source of influence.

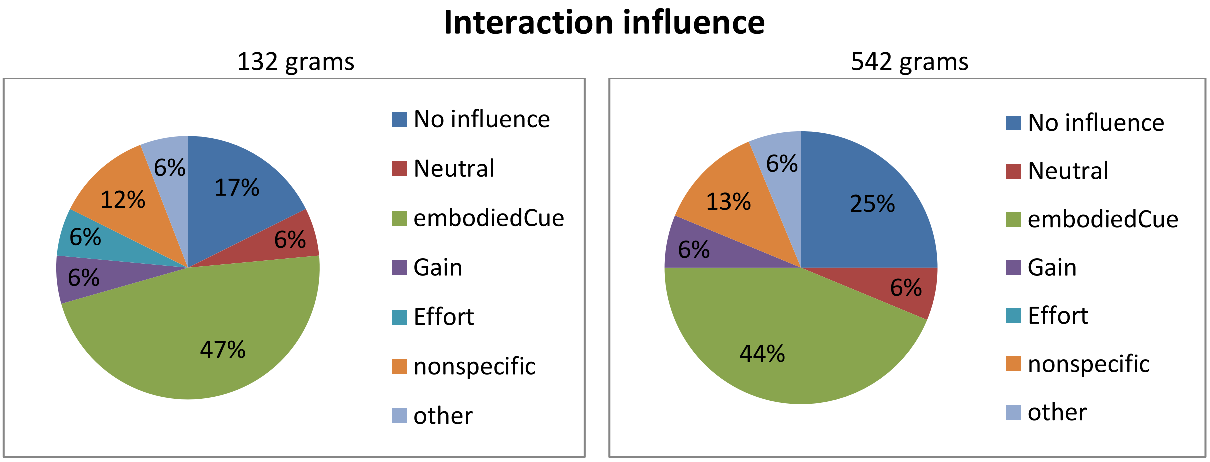
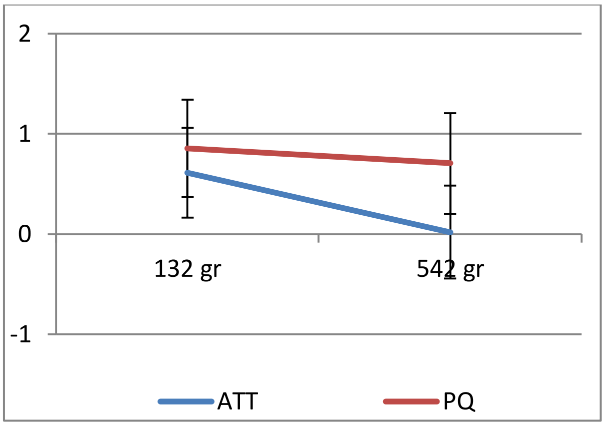


Figure 19 Overview of subjective rating for influence of interaction on estimations per condition

One important alteration to the post-experimental questionnaire for Experiment 2 was the inclusion of a fatigue question. As it was expected that using a heavier device would lead to more fatigue for the users, this question was introduced to function as a manipulation check. Based on the Borg CR-10 scale (see Appendix E) participants filled out an appropriate value based on a provided ratio scale to describe their physical fatigue. Comparing these values between the weight conditions showed that participants who received the heavier controller reported being more fatigued (*F(1, 31) = 6.65, p = .015,* η^2 *= .177*). Mapping the means for both groups back on the scale, the low weight group (*M = 2.06, SE = 0.44*) has indicated low to medium levels of fatigue while the high weight group (*M = 3.69, SE = 0.45*) reported medium to somewhat high levels of fatigue.

Finally, the user experience was measured using the AttrakDiff2 questionnaire (Hassenzahl, 2007a). Unlike the first experiment, however, the focus was not placed on the apparatus but on the interaction itself. The scores for the items that made up each dimension were averaged to form their respective dimension score. An ANOVA analysis was then done to consider whether differences between these scores existed following the weight manipulation. For three of the dimensions no significant differences found: Pragmatic Quality (*F(1, 31) = 0.20, p = .661,* η^2 *= .006*), Hedonic Quality-Identification (*F(1, 31) = 3.54, p = .069,* η^2 *= .102*) and Hedonic Quality-Stimulation (*F(1, 31) = 0.74, p = .397,* η^2 *= .023*). In terms of general Attractiveness, however, the lower weight controller was seen as significantly more attractive than the higher weight controller (*F(1, 31) = 5.97, p = .020,* η^2 *= .162*).



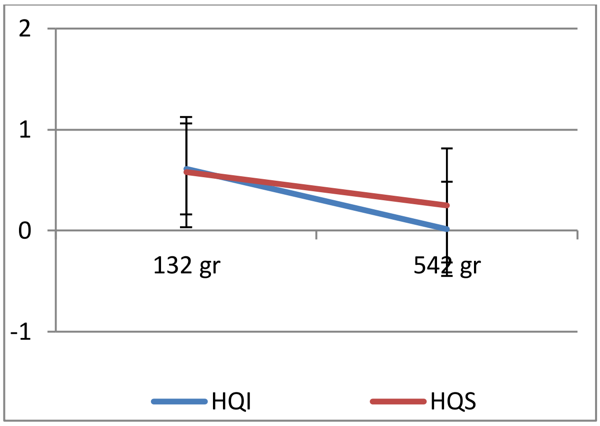


Figure 20 User experience dimension scores (with 95% CI). Left: Attractiveness & Pragmatic Quality; Right: Hedonic Quality-Identification & Hedonic Quality-Stimulation

The results from the questionnaire do not support the second and third hypothesis. While the pragmatic quality dimension remains nearly equal, both hedonic quality dimensions show a non-significant opposite trend. Despite not being significant, the effect size of the HQI is fairly large.

4.3 Discussion

The main aim of the second experiment was to retest the effort hypotheses using a different setup that was based on van Beurden (2013) experiment 9. A key change following the new setup was that a different type of device with a different interaction method was required. The flexibility of the new setup allowed for the development and use of two custom tailored tracker devices and therefore the application of a stronger manipulation. The methodological differences also led to changes in the post-experimental questionnaire. The questionnaire was extended to incorporate a manipulation check and the full Attrakdiff2 questionnaire (Hassenzahl, 2007a).

The first analysis was done to form an overall view of the estimations. As in van Beurden (2013) it was found that participants tended to underestimate the distances in the VE for all types of presentations. Hereby the underestimations were especially strong for the lower egocentric distances. A possible explanation may be found in the limited field of view (FOV) due to the HMD in combination with participant looking behavior. The limited FOV caused that, when looking straight ahead in the VE, the end of the table, the place where participants were located, was not visible. However, when looking down, due to the head tracking, participants would be able to see where the table ended and their virtual representation was located. From observing participants' looking behavior it was noticed that only few participants used the ability to fully look down despite not being restricted in doing so. In the past it has been suggested that information acquired by head movements can reduce underestimations in VE (see Campos, Nusseck, Wallraven, Mohler & Bülthoff, 2007 for a discussion). The lack of environmental exploration could thus pose one possible explanation for the underestimations.

Considering the main hypothesis, for the current setup, the results of the exocentric and egocentric distance estimation task do not provide grounds on which the first hypothesis can be accepted. No significant differences were found between the weight groups, nor was there a trend that indicated that the hypothesis may be found for a larger sample. There are several possible reasons for the lack of findings, the most important of which should be sought in the methodology.

As it was believed that the methodology may have played part in the lack of findings several methodological aspects were analyzed. First, the setting was intended to prevent response strategy usage and it could be established that it was successful at its purpose. All participant behavior in the VE could be monitored on a display that was visible for the experimenter and little to no irregular behavior that could indicate response strategy usage was observed. Another problem in the previous experiment was the manipulation strength. The post-experimental questionnaire was extended with a physical fatigue question which, in addition to a question about apparatus weight, was included to assess the success of the manipulation. The apparatus weight question in the first part of the post-experimental questionnaire showed that participants rated the weight of the heavy controller as higher. Second, the fatigue question indicated that participants who received the heavier controller were significantly more fatigued after the experiment. Despite that a difference was likely felt, it is possible that the extent to which it was experienced in terms of effort was insufficient to trigger an effort effect. In this matter the fatigue results should particularly not be over relied upon as the measurement was limited in two important ways. First, since the question was preceded by the funneled question series regarding the experimental goal, influence of these prior questions cannot be ruled out. Second, the questionnaire was filled out electronically and because of wrong browser settings, history was remembered for some text fields and some results filled out by previous participants may have been exposed in some cases. The belief that the manipulation may have been too small can also be supported by the fact that nearly none of the participants noted weight to be out of the ordinary. Although, this last observation may also be explained by other factors such as the fact that participants did not see the device, or had less experience with motion tracking devices in general and therefore lacked a strongly established beliefs on the weight of such a device.

Another possible explanation could be found in the validity of Effort Theory. Given that none of the participants found out about the experimental intentions and the fact that the heavier apparatus was both experienced as heavier and more fatiguing, the null finding could be seen as problematic for an effort account. However, based on the uncertainties involving the manipulation strength, there appears to be little grounds to assume this was the case.

The second and third hypothesis were directed at determining user experience of the interaction on a pragmatic (‘do’) and hedonic (‘be’) level. It was expected that pragmatic quality would decrease with increases in weight while hedonic quality would increase. However, the questionnaire data showed neither of these effects. The pragmatic dimension remained near constant across the weight conditions while the hedonic quality dimensions showed downward trends with weight increases. Particularly the Hedonic Quality-Identification trend seemed to indicate that participants identified less with the heavier device interaction.

These results are surprising and explanations should initially be sought in the limitations of the current experiment and differences between the two experiments. The null finding for the pragmatic quality dimension could be seen as more evidence for the weak manipulation notion. When the difference in effort required to perform a task is too small, no significant pragmatic quality difference between the weight conditions should be expected. The results for the hedonic quality dimensions are more difficult to explain. One possibility could lie in the difference between the devices that were used and the meaning that weight has for each of these devices. While heavier computer mice are sometimes marketed as special (e.g. as made for gaming), weight variations in motion controllers appear to be uncommon and may not hold a similar meaning for the user. In other words, weight could have a conceptually different meaning between these devices. Another possible explanation for the unexpected results lies in the different focus on interaction. While a link between interaction device user experience and user experience of the interaction mediated by such a device was expected, it could be that properties of the interaction device are differently weighted when interaction itself is rated.

A final limitation for the current experiment was its limited sample size which in combination with a violation of the normality assumption for the distance task data formed a problem for the parametric analysis. These issues were reflected in the kurtosis of the measurements and the issues differed per measurement and condition, some were leptokurtic while others were platykurtic. While robust solutions have been coined for such problems (Wilcox, 2005), these solutions have yet to be implemented by default in commonly used statistical software packages and can be somewhat cumbersome. In light of the small effect sizes for the manipulation, and because the statistical issues do not appear to have a systematic character, the currently applied analysis is maintained as a good estimation of what the results would be like using a robust analysis.

5. General Discussion

The current thesis aimed to investigate a possible influence of effortful interaction on visual perception and user experience. These questions were studied by conducting conceptual replications of Experiments 8 and 9 by van Beurden (2013). Based on the premises of the effort-calibration hybrid position by Woods et al. (2009), both the experimental context of the before mentioned experiment and the Effort Theory literature were considered to determine an Effort Theory hypothesis. The resulting hypothesis was that there would be an effect of effortful interaction on distance perception. An influence of effortful interaction on user experience was investigated using a reductionist user experience model (Hassenzahl, 2007a). Two user experience hypotheses were derived, one for an influence of effortful interaction on pragmatic quality, and one for such an influence on hedonic quality. The latter was based on findings in a pilot study of the first experiment which indicated that user experience may be affected by an effort induction. Hence it was predicted that introducing effort in the interaction reduces perceived pragmatic quality, and would lead to increases in perceived hedonic quality.

This chapter will discuss the findings from both experiments in relation to each other, their implications for the interpretation of earlier work, and will propose future research possibilities.

5.1 Influence of Effort on Distance Perception

The first goal of the current study was to investigate whether an effortful interaction could influence both exocentric and egocentric distance perception in mediated environments. Before discussing the results it is important to first reflect upon an important part of this goal, the validity of the results. Because the current study investigated an Effort Theory explanation for an earlier finding (van Beurden, 2013), and Effort Theory has been criticized due to possible demand characteristics (e.g. Durgin et al., 2009), measures had to be taken to ensure that any possible effect would not be due to demand characteristics. In both experiments a post-experimental questionnaire was conducted (see Appendix B for the version used in Experiment 1). Part of this questionnaire was a series of questions that followed a funneled procedure which, over the course of the questions, would provide small hints about the intentions that would help determine at what point during the questionnaire participant responses started to show suspicions of the research hypothesis. The questionnaire that was used was found to be highly effective for this purpose. During the debriefing of the first experiment participants expressed no beliefs about the involvement of the apparatus in the experimental design, however, indicated that the small hints had informed them on its relevance. Based on the questionnaire results it could be established that for both experiments few participants displayed suspicions about the main hypothesis and were thus naïve about the experimental aim.

The first experiment employed a standard desktop setup and manipulated the mouse weight beween-subjects. In this experiment no differences were found between the weight groups. Because several limitations of the setup were identified including the manipulation strength and response strategy use by the participants, and to further investigate van Beurden’s results, a second experiment was planned. In the second experiment, a Head Mounted Display (HMD) was used which constrained response strategy usage due to the lack of bezel which could be used as a reference and the fact that the screen could no longer be touched. Moreover, a different and heavier interaction device was used to create a stronger effort induction which was also predicted to be experienced to a larger extent. The main reason for this expectation was that the controller would no longer be supported by a horizontal surface, which in the past has been suggested to be more fatiguing (Zeng et al., 2012). Thus, many of the limitations of the first experiment were believed to have been eliminated or at least reduced. In contrast to expectations however, the results of Experiment 2 again showed no effect of weight on visual perception. Hence, based on the findings from both experiments, hypothesis 1 cannot be accepted.

In the Effort Theory literature null-findings have previously been reported (e.g. Woods et al., 2009) and are sometimes seen as evidence against the validity of the theory (Durgin et al., 2009). At first glance the results may thus appear problematic for Effort Theory. This belief may be strengthened by the lack of suspicions about the experimental aim which could be proposed as an reason for the null-findings. While such an explanation cannot entirely be ruled out, it perhaps too easily assumes that the effort differences were sufficiently strong. However, the measurements used to test the experienced effort are limited in predicting to what extent effort was experienced. The weight question in particular should not be directly linked to experienced effort as weight ratings may have been based on the device context i.e. participants believed that the apparatus was heavy within the scope of being a member of a certain device category (e.g. mouse). To resolve this limitation, a fatigue item was added in the second experiment. However, this measure may not have been free from bias due to questionnaire ordering and a technical error. Beyond a lack of evidence for sufficiently experienced effort, there may even be evidence against the belief that experienced effort strongly differed between conditions. It seems likely that when interaction becomes truly effortful due to the apparatus weight, a larger proportion of the participants would indicate earlier in the questionnaire that the device was heavy. However, across both experiments in the early questions few indicated that this was the case.

Despite having established that the manipulation strength may have been too small to cause an effect, there is still value in the results from the two experiments. Relating the results to van Beurden (2013) provides insight into the possible mechanism underlying the findings in that study. A key question is whether the evoked effort that stemmed from the gain manipulation can be compared to the levels of effort required due to the weight manipulation in the current work. Unfortunately, with the currently available data, an answer to this question cannot be given unequivocally. Drawing a comparison between the employed manipulations is complicated partly because no common effort measurements have been conducted between all experiments. Nevertheless, in light of the current results it seems unlikely that the difference between effort levels was of such magnitudes that an effort mechanism could explain the gain manipulation findings. A partial contribution of effort in the gain study findings, however, cannot entirely be ruled out.

5.2 Influence of Effort on User Experience

The second goal for the current thesis was to determine whether effortful interaction could affect user experience. While both experiments focused on user experience, there was one important difference. The first study investigated user experience of the apparatus while the second was directed at user experience of the interaction. The hypotheses formed related to three dimensions of user experience Pragmatic Quality, Hedonic Quality-Identification and Hedonic Quality-Stimulation (Hassenzahl, 2007a). Since in Experiment 1 only a subset of the Attrakdiff2 questionnaire was used and items lacked sufficient correlation, the findings from this study were analyzed on a per item basis. The items did, however, show some interesting and related trends that matched the hypotheses. The decrease of pragmatic quality items (complicated-simple, impractical-practical & unpredictable-predictable) with weight indicated that participants believed that the device was less capable of assisting in fulfillment of the task at hand. On a hedonic level no indications were found that participants could identify more with the heavier device. However, participants may have felt more stimulated. In the second experiment the user experience approach was somewhat different due to a focus on the interaction. Because it has previously been suggested that physical properties of the device may impact user experience of the interaction (Jordan, 2000), and it seems likely that in the first experiment the device user experience was based upon physical characteristics, it was expected that similar results would be found. Interestingly this was not the case as neither the pragmatic nor the hedonic dimensions showed the expected effects.

One possible explanation for the discrepant findings concerns the view that participants have of the device. It seems likely that every participant has worked with a mouse before or even works with one on a regular basis. This may not be so much the case for a motion controller. In the former case it appears more likely that the participant is aware of the range of common weights for the device or at least is familiar with the approximate weight of one exemplar. A deviation from the existing expectations may therefore be more salient. Salience of the weight property may in the first experiment have caused the relatively larger effect sizes for the pragmatic quality dimension items despite that the actual weight difference was smaller than in the second experiment. Differences in familiarity with and the availability of held beliefs between the devices may also explain the different results on the hedonic dimensions. From the data it could be found that the weight differences in both experiments were significant. The meaning of a higher weight may, however, vary between the devices. While computer mice may be seen as special when they are heavier, this belief does not necessarily transfer to interaction with motion controllers. Indeed, weight adjustable computer mice are commercially available and marketed for special audiences (e.g. video game players). For motion controllers, however, weight does not appear to be a common differentiating factor between products. This difference may explain the stimulation dimension result in the first experiment.

Another account that could be coined for both the unexpected contrasts in pragmatic and hedonic quality could be the shift in focus from the device to the interaction itself. It was expected that links between experience of the interaction and experience of the device existed due to earlier indications of the inclusion of physical apparatus properties in user experience of interaction. However, the interplay between device user experience and broader interaction user experience may not be straightforward as the latter consists of more factors of which the relative weighting remains unknown. Finally, procedural differences such as the lack of visibility of the apparatus between the experiments may have contributed to the mixed findings. The variety in possible explanations shows that more research on the effect of effort on user experience seems desirable. Particularly given that both of the current experiments were limited in one important way.

The user experience findings are limited in one important way. In both studies, the order of the experimental tasks was the same. The funneled procedure questions directly followed the distance estimation tasks. This sequence was established because introducing intermittent tasks between the distance estimation task and the funneled procedure questions could lead to a distorted image of the participant suspicion at the time of the distance estimation task. While this approach rules out external influence on the funneled procedure question, influence in the opposite direction should also be considered. Indeed, the funneled procedure questions could have affected the responses on the user experience questions. This is especially a concern given that the first part of the questionnaire explicitly asks about factors pertaining to the interaction. Participants may thus have become aware of the relevance of these factors in the study which in turn could have influenced the user experience results.

5.3 Future Research

Based on the results of the current thesis several possible research future research directions could be considered.

While no effort effects were found for the methodologies as used in both experiments, it would be interesting to consider whether under different conditions effort effects can be found. While some methodological changes should be considered, the setup as used in the second experiment could still be used since it has proven unaffected by response strategies and the fact that there may be more headroom for weight manipulations without raising suspicion. An obvious continuation with this setup could thus be to further increase either the weight of the apparatus or the duration of use to see whether effects can be found. Alternatively, different interaction methods could be explored that feature a stronger embodied character and provide less physical support. Especially, free-handed gesture interaction, which can be more fatiguing than device based interaction (van Beurden, 2013), seems interesting to investigate due to its potential practical relevance. Fatigue based effort manipulations could be employed in which participants play a gesture based game or do a performance based task for a longer duration before measuring visual perception in a distance estimation task. Since participants should not be aware of the manipulation the tasks could be explained as separate experiments.

In a later stage when any of the above proposed methodologies would lead to found effects, research efforts could investigate the effect’s robustness and origin. Measurement types could be varied or additional indirect measures could be included such that the principle of converging measurements can be relied on (Loomis & Philbeck, 2008). Moreover, additional research efforts could investigate whether the effect can be attributed to visual perception or whether it is memory based. In the setups employed in the current experiment only one point in virtual space was displayed at a time. A stronger case for an effort effect on visual perception rather than memory could be made when similar effects also occur with both points visible at the same time.

The findings in the current thesis not only have implications for the interpretation of van Beurden’s (2013) results but also for future research directions. One of the limitations in that study has been that no post-experimental assessment of bias was conducted. Due to the nature of the measurements conducted it may be interesting to replicate the experiments and include a bias assessment. In this way, upon successful replication it can be established whether participants were unaware of the gain manipulation or simply complied to the experimental expectations. Additionally, because an effort account seems unlikely to explain the full effect reported in the before mentioned study, other accounts that have been proposed should primarily receive attention in future studies that aim to establish the cause of the effect.

The second goal of this thesis was to investigate whether effort would impact user experience. From the conflicting findings new questions have risen. Among others, these questions revolve around the apparatus weight which could have a different meaning depending on device type, and the extent to which user experience of a physical apparatus influences user experience of the interaction with that apparatus. Both could be explored in future research endeavors. To prevent any order limitations as identified in the current experiments, future studies in this direction are best to be detached from Effort Theory related studies.

6. Concluding Remarks

This thesis has investigated the impact of effort within human computer interaction by considering possible effects on visual perception and user experience. Hereby it is one of the first that has considered the possibility of an Effort Theory mechanism within this specific scope. In two experiments it was shown that using differently weighted devices under the conditions set by both experiments, did not lead to levels of effort exertion that cause an influence on visual perception. It cannot, however, be ruled out that under different conditions (e.g. stronger effort manipulations) effects can be found. Therefore, the results arguably do not impact the status quo of Effort Theory itself. Despite the apparent lack of Effort Theory related findings, the results appear theoretically relevant as they are informative about the mechanism functional in van Beurden (2013). Because the levels of effort exerted in the current experiments and in van Beurden are believed to be relatively similar, a strictly effort based explanation for the gain findings in the latter appears to be unlikely.

The user experience questions had a somewhat exploratory character and due to a combination of mixed findings and limitations it has been difficult to draw firm conclusions from the results. However, new and interesting questions regarding the influence of effort on user experience have been raised that, when answered, could shine more light on the interplay between device properties such as weight and general user experience.

Overall, studying effects of effortful interaction on the user both in terms of perception and user experience of a system is an interesting research direction that deserves more attention. The results and corresponding reflections in the current thesis could therefore lead the way into many more studies to come.

7. References

Abowd, G. D. & Mynatt, E. D. (2000). Charting past, present, and future research in ubiquitous computing. *ACM Transactions on Computer-Human Interaction (TOCHI)*, *7*, 29-58.

Bakeman, R. (2005). Recommended effect size statistics for repeated measures designs. *Behavior Research Methods*, *37*, 379-384.

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, *59*, 617-645.

Baudel, T., & Beaudouin-Lafon, M. (1993). Charade: remote control of objects using free-hand gestures. *Communications of the ACM - Special issue on computer augmented environments: back to the real world*, *36*, 28-35.

Bhalla, M., & Proffitt, D. R. (1999). Visual-motor recalibration in geographical slant perception. *Journal of Experimental Psychology: Human Perception and Performance, 25,* 1076–96.

Borg, G. (1990). Psychophysical scaling with applications in physical work and the perception of exertion. *Scandinavian Journal of Work, Environment & Health*, *16,* 55-58.

Cabral, M. C., Morimoto, C. H., & Zuffo, M. K. (2005, October). On the usability of gesture interfaces in virtual reality environments. In *Proceedings of the 2005 Latin American conference on Human-computer interaction* (pp. 100-108). New York, NY: ACM.

Carmines, E. & Zeller, R. (1979). *Reliability and Validity Assessment. Sage Paper Series on Quantitative Applications No. 07-017*. Beverly Hills, CA: Sage.

Campos, J., Nusseck, H. G., Wallraven, C., Mohler, B. J., & Bülthoff, H. H. (2007). Visualization and (mis)perceptions in virtual reality. *Tagungsband, 10,* 10-14.

Cooper, A. D., Sterling, C. P., Bacon, M. P. & Bridgeman, B. (2012) Does action affect perception or memory? *Vision Research, 62,* 235-240.

Creem-Regehr, S. H., Gooch, A. A., Sahm, C. S., & Thompson, W. B. (2004). Perceiving virtual geographical slant: Action influences perception. *Journal of Experimental Psychology: Human Perception and Performance*, *30*, 811-821.

Durgin, F. H., Baird, J. A., Greenburg, M., Russell, R., Shaughnessy, K., & Waymouth, S. (2009). Who is being deceived? The experimental demands of wearing a backpack. *Psychonomic Bulletin & Review*, *16*, 964-969.

Durgin, F. H., Klein, B., Spiegel, A., Strawser, C. J., & Williams, M. (2012). The social psychology of perception experiments: Hills, backpacks, glucose, and the problem of generalizability. *Journal of Experimental Psychology: Human Perception and Performance*, *38*, 1582-1595.

Durgin, F. H., & Li, Z. (2010). Controlled interaction: Strategies for using virtual reality to study perception. *Behavior Research Methods*, *42*, 414-420.

Firestone, C. (2013). How “paternalistic” is spatial perception? Why wearing a heavy backpack doesn’t—and couldn’t—make hills look steeper. *Perspectives on Psychological Science*, *8*, 455-473.

Gibson, J. J. (1986). *The ecological approach to visual perception*. St. Joseph, Missouri: Psychology Press.

Gogel, W. C. (1976). An indirect method of measuring perceived distance from familiar size. *Perception & Psychophysics*, *20*, 419-429.

Gordon, I. E. (2004). *Theories of visual perception* (3rd ed.). New York, NY: Psychology Press.

Greeno, J. G. (1994). Gibson’s affordances. *Psychological Review*, *101*, 336–342.

Haber, R. N., & Levin, C. A. (2001). The independence of size perception and distance perception. *Perception & Psychophysics*, *63*, 1140-1152.

Hamburger, K., Hansen, T. (2010). Analysis of individual variations in the classical horizontal – vertical illusion. *Attention, Perception, & Psychophysics, 72,* 1045–1052.

Hassenzahl, M. (2007a). AttrakDiff(tm). Retrieved from http://attrakdiff.de/index-en.html.

Hassenzahl, M. (2007b). The Hedonic/Pragmatic Model of User Experience. In Law, E., Vermeeren, A., Hassenzahl, M., & Blythe, M. (Eds.) *Towards a UX Manifesto – Proceedings of a cost294-affiliated workshop on HCI 2008* (pp. 10-14). Lancester, UK.

Hassenzahl, M. (2004). The thing and I: Understanding the relationship between user and product. In Blythe, M.A., Overbeeke, K., Monk, A.F. and Wright, P. C. (Eds.), *Funology: From Usability to Enjoyment*. New York, NY: Kluwer Academic Publishers.

Hurkmans, H. L., Ribbers, G. M., Streur-Kranenburg, M. F., Stam, H. J., & van den Berg-Emons, R. J. (2011). Energy expenditure in chronic stroke patients playing Wii Sports: a pilot study. *Journal of Neuroengineering and Rehabilitation*, *8*, 1-7.

Hutchison, J. J., & Loomis, J. M. (2006a). Does energy expenditure affect the perception of egocentric distance? A failure to replicate Experiment 1 of Proffitt, Stefanucci, Banton, and Epstein (2003). *The Spanish Journal of Psychology*, *9*, 332-339.

Hutchison, J. J., & Loomis, J. M. (2006b). Reply to Proffitt, Stefanucci, Banton, and Epstein. *The Spanish Journal of Psychology, 9,* 343–345.

Jordan, P. W. (2000). *Designing Pleasurable Products: An introduction to the new human factors*. London: Taylor & Francis.

Kim, Y., Lee, G. A., Jo, D., Yang, U., Kim, G., & Park, J. (2011, January). Analysis on virtual interaction-induced fatigue and difficulty in manipulation for interactive 3D gaming console. In *Consumer Electronics (ICCE), 2011 IEEE International Conference on* (pp. 269-270). IEEE.

Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology, 49,* 764–766.

Loomis, J.M., & Philbeck, J.W. (2008). Measuring perception with spatial updating and action. In R.L. Klatzky, M. Behrmann, & B. MacWhinney (Eds.), *Embodiment, ego-space, and action* (pp. 1–43). Mahwah, NJ: Erlbaum.

Lyons, L., Lopez Silva, B. L., Moher, T., Jimenez Pazmino, P., & Slattery, B. (2013, June). Feel the burn: exploring design parameters for effortful interaction for educational games. In *Proceedings of the 12th International Conference on Interaction Design and Children* (pp. 400-403). New York, NY: ACM.

Lyons, L., Slattery, B., Jimenez Pazmino, P., & Moher, T. (2012, February). Don't forget about the sweat: Effortful embodied interaction in support of learning. In *TEI '12 Proceedings of the Sixth International Conference on Tangible, Embedded and Embodied Interaction* (pp. 77–84). New York, NY: ACM.

Mahlke, S. (2007). *User experience of interaction with technical systems*. (Doctoral dissertation). Retrieved from http://www.zmms.tu-berlin.de/~sma/ressources/theses/.

Michaels, C. F. (2003). Affordances: Four points of debate. *Ecological Psychology*, *15*, 135-148.

Michaels, C. F., & Carello, C. (1981). *Direct perception* (pp. 1-208). Englewood Cliffs, NJ: Prentice-Hall.

Norman, J. (2002). Two visual systems and two theories of perception: An attempt to reconcile the constructivist and ecological approaches. *The Behavioral and Brain Sciences*, *25*, 73–96.

Padoy, N., & Hager, G. (2011). Gesture-based Manipulation of a Surgical Tool using a Kinect (JHU, 2011). *Research homepage of Nicolas Padoy*. Retrieved November 13, 2013, from http://www.cs.jhu.edu/~padoy/doku.php/videos.

Peng, W., Lin, J. H., & Crouse, J. (2011). Is playing exergames really exercising? A meta-analysis of energy expenditure in active video games. *Cyberpsychology, Behavior, and Social Networking*, *14*, 681-688.

Philbeck, J. W., Woods, A. J., Kontra, C., & Zdenkova, P. (2010). A comparison of blindpulling and blindwalking as measures of perceived absolute distance. *Behavior Research Methods, 42,* 148–160.

Proffitt, D. R. (2006). Embodied perception and the economy of action. *Perspectives on Psychological Science*, *1*, 110–122.

Proffitt, D. R. (2013). An embodied approach to perception: by what units are visual perceptions scaled? *Perspectives on Psychological Science, 8,* 474–483.

Proffitt, D. R., Bhalla, M., Gossweiler, R., and Midgett, J. (1995). Perceiving geographical slant. *Psychonomic Bulletin & Review, 2,* 409–428.

Proffitt, D. R., Stefanucci, J., Banton, T., & Epstein, W. (2003). The role of effort in perceiving distance. *Psychological Science*, *14*, 106-112.

Proffitt, D. R., Stefanucci, J., Banton, T., & Epstein, W. (2006). Reply to Hutchison and Loomis. *The Spanish Journal of Psychology, 9,* 340–342.

Ruppert, G. C. S., Reis, L. O., Amorim, P. H. J., de Moraes, T. F., & da Silva, J. V. L. (2012). Touchless gesture user interface for interactive image visualization in urological surgery. *World Journal of Urology*, *30*, 687-691.

Schnall, S., Zadra, J. R., & Proffitt, D. R. (2010). Direct evidence for the economy of action: Glucose and the perception of geographical slant. *Perception, 39,* 464-482.

Slattery, B., Lyons, L., Lopez Silva, B., & Jimenez Pazmino, P. (2013). Extending the reach of Embodied interaction in informal spaces. Poster presented at *the 10th international conference on Computer Supported Collaborative Learning (CSCL 2013)*. Madison, WI. Retrieved from http://bslatt2.people.uic.edu/pdf/slattery-et-cscl-2013.pdf

Thompson, J. G., & Schiffman, H. R. (1974). The influence of figure size and orientation on the magnitude of the horizontal-vertical illusion. *Acta Psychologica*, *38*, 413-420.

Van Beurden, M. H. P. H. (2013). *Interaction in Depth*. (Doctoral dissertation). Retrieved from http://repository.tue.nl/760050.

Wahlström J. (2001). Physical load in computer mouse work: working technique, sex and stress aspects. In Marklund S, Meding B, Rosén G, Tronqvist EW (Eds.), *Arbete och hälsa* (pp 1–34). Stockholm, Sweden: National Institute for Working Life.

Wilcox, R. R. (2005). *Introduction to robust estimation and hypothesis testing* (3rd ed.). Elsevier.

Witt, J. K. (2011). Action's effect on perception. *Current Directions in Psychological Science*, *20*, 201-206.

Witt, J. K., & Dorsch, T. E. (2009). Kicking to bigger uprights: Field goal kicking performance influences perceived size. *Perception*, *38,* 1328-1340.

Witt, J. K., & Proffitt, D. R. (2005). See the Ball, Hit the Ball Apparent Ball Size Is Correlated With Batting Average. *Psychological Science*, *16*, 937-938.

Witt, J.K., Proffitt, D.R., & Epstein, W. (2004). Perceiving distance: A role of effort and intent. *Perception, 33,* 577–590.

Witt, J. K., Proffitt, D. R., & Epstein, W. (2010). When and how are spatial perceptions scaled? *Journal of Experimental Psychology: Human Perception and Performance*, *36*, 1153-1160.

Woods, A. J., Philbeck, J. W., & Danoff, J. V. (2009). The various perceptions of distance: an alternative view of how effort affects distance judgments. *Journal of Experimental Psychology: Human Perception and Performance*, *35*, 1104-1117.

Yoo, B., Han, J. J., Choi, C., Yi, K., Suh, S., Park, D., & Kim, C. (2010, April). 3D user interface combining gaze and hand gestures for large-scale display. In *CHI'10 Extended Abstracts on Human Factors in Computing Systems* (pp. 3709-3714). ACM.

Zeng, X., Hedge, A., & Guimbretiere, F. (2012, September). Fitts' Law in 3D Space with Coordinated Hand Movements. *In Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 56, No. 1, pp. 990-994). SAGE Publications.

Appendices

Appendix A: Attrakdiff2 Dutch translation

	German	English	Dutch
PQ 1	technisch- menschlich	technical – human	technisch – menselijk
2	kompliziert - einfach	complicated - simple	ingewikkeld - simpel
3	unpraktisch - praktisch	impractical - practical	onpraktisch - praktisch
4	umständlich – direkt	cumbersome – straightforward	omslachtig - direct
5	unberechenbar – voraussagbar	unpredictable - predictable	onvoorspelbaar - voorspelbaar
6	verwirrend – übersichtlich	confusing – clearly structured	verwarrend - overzichtelijk
7	widerspenstig - handhabbar	unruly - manageable	onhandelbaar - handelbaar
HQI 8	isolierend – verbindend	isolating – connective	isolerend - verbindend
9	laienhaft - fachmännisch	unprofessional – professional	onprofessioneel – professioneel
10	stillos – stilvoll	tacky - stylish	stijlloos – stijlvol
11	minderwertig – wertvoll	cheap - premium	goedkoop - waardevol
12	ausgrenzend - einbeziehend	alienating – integrating	vervreemdend - integrerend
13	trennt mich – bringt mich näher	seperates me – brings me closer	scheidt me af – brengt me nader
14	nicht vorzeigbar – vorzeigbar	unpresentable – presentable	niet toonbaar - toonbaar
HQS 15	konventionell – originell	conventional - inventive	conventioneel - origineel
16	phantasielos – kreativ	unimaginative - creative	fantasieloos - creatief
17	vorsichtig – mutig	cautious - bold	voorzichtig - moedig
18	konservativ – innovative	conservative -innovative	conservatief - innovatief
19	lahm – fesselnd	dull – captivating	saai – boeiend
20	harmlos - herausfordernd	undemanding - challenging	weinig eisend - Uitdagend
21	herkommlich – neuartig	ordinary - novel	gewoon – ongebruikelijk
ATT 22	unangenehm – angenehm	unpleasant - pleasant	onplezierig - plezierig
23	hässlich - schön	ugly - attractive	lelijk – mooi
24	unsympatisch – sympatisch	disagreeable - likeable	onaangenaam – aangenaam
25	zurückweisend – einladend	rejecting - inviting	afwijzend - uitnodigend
26	slecht - gut	bad - good	slecht - goed
27	abstossend – anziehend	repelling - appealing	afstotend - aantrekkelijk
28	entmutigend - motivierend	discouraging - motivating	ontmoedigend – motiverend

Appendix B : Experiment 1 Questionnaire

Vragenlijst

Comp. NR: PPNR:

Bedankt voor het meedoen aan het experiment en het invullen van deze vragenlijst. U wordt verzocht de volgende vragen achtereenvolgend te beantwoorden. Ga niet terug te gaan naar een vorige vraag of pagina, maar beantwoord de vragen van voor naar achteren.

Wat is uw leeftijd? _____ jaar

Wat is uw geslacht? O Vrouw O Man

Uw ervaring met de experimentele taak

De volgende serie vragen heeft betrekking op uw ervaring met de experimentele taak. Sommige van deze vragen bevatten een open deel. Omdat we benieuwd zijn naar hoe uw antwoord tot stand is gekomen wordt u verzocht hier uw antwoord kort maar duidelijk toe te lichten.

1. Waar denkt u dat dit experiment over ging?
Overweeg zaken die mogelijk opvielen tijdens het experiment. Als u geen idee heeft schrijf dan: "Ik weet het niet"

Geef uw mening over de volgende stelling

2. "De interactie met de muis ging zoals ik vooraf had verwacht"

Zeer mee oneens	O 1	O 2	O 3	O 4	O 5	O 6	O 7	Zeer mee eens

Leg hieronder uw antwoord uit:

Beantwoord de volgende vragen

3. Hoe heeft u het gewicht van de muis ervaren tijdens het experiment?

Zeer laag	O 1	O 2	O 3	O 4	O 5	O 6	O 7	Zeer hoog

4. Hoe heeft u de gain van de muis ervaren tijdens het experiment?

Gain beschrijft de relatie tussen fysieke beweging van de muis en beweging van de cursor op het scherm. Een lage gain betekent een grote beweging van de arm en een hoge gain een kleine beweging van de arm voor dezelfde cursor verplaatsing.

Zeer laag	O 1	O 2	O 3	O 4	O 5	O 6	O 7	Zeer hoog

5. Hoe waarschijnlijk acht u het dat de interactie met de muis uw afstandsschattingen heeft beïnvloedt?

Zeer on-waarschijnlijk	O 1	O 2	O 3	O 4	O 5	O 6	O 7	Zeer waarschijnlijk

Licht hieronder uw antwoord toe:

Uw mening over de muis die u tijdens het experiment heeft gebruikt

Deze pagina bevat een set met woordparen. Geef met behulp van deze woordparen aan wat u de meest passende beschrijving vindt voor de muis die u heeft gebruikt tijdens het experiment.

Aangenaam	O 1	O 2	O 3	O 4	O 5	O 6	O 7	Onaangenaam
Origineel	O 1	O 2	O 3	O 4	O 5	O 6	O 7	Conventioneel
Ingewikkeld	O 1	O 2	O 3	O 4	O 5	O 6	O 7	Simpel
Lelijk	O 1	O 2	O 3	O 4	O 5	O 6	O 7	Mooi
Praktisch	O 1	O 2	O 3	O 4	O 5	O 6	O 7	Onpraktisch
Stijlvol	O 1	O 2	O 3	O 4	O 5	O 6	O 7	Stijlloos
Voorspelbaar	O 1	O 2	O 3	O 4	O 5	O 6	O 7	Onvoorspelbaar
Goedkoop	O 1	O 2	O 3	O 4	O 5	O 6	O 7	Waardevol
Goed	O 1	O 2	O 3	O 4	O 5	O 6	O 7	Slecht
Afstotend	O 1	O 2	O 3	O 4	O 5	O 6	O 7	Aantrekkelijk
Innovatief	O 1	O 2	O 3	O 4	O 5	O 6	O 7	Conservatief

Motiverend	O 1	O 2	O 3	O 4	O 5	O 6	O 7	Ontmoedigend
Ongebruikelijk	O 1	O 2	O 3	O 4	O 5	O 6	O 7	Gewoon

Wat is uw geprefereerde hand?	O Linker hand	O Rechter hand
Met welke hand heeft u de muis gebruikt?	O Linker hand	O Rechter hand

---

Was u al bekend met deze muis

Bijvoorbeeld: u gebruikt(e) deze muis thuis

O Ja

O Nee

O Ik weet het niet

Dit is het einde van de vragenlijst, bedankt voor uw deelname.

Appendix C : Overview of Questionnaire Coding

Experiment 1

Question 1 - Direct Experimental Suspicions

Code	Full	Explanation
horvert	Horizontal-Vertical difference	Participants expressed the belief that there were differences between perception of horizontally and vertically presented distances or that the goal of the experiment was to measure whether this was the case.
generic	Generic	Participant did not report a specific hypothesis or did not explain in a comprehensible way.
posonscreen	Position on screen	Participant suspected that the alignment of the two buttons was part of the experimental manipulation
dontknow	Don't know	Participant reported not knowing the experimental aim.
mouseweight	Mouse weight	Participant suspected the mouse weight to be related to the experimental goal
other	Other	Various other explanations

Question 2 – Mouse interaction as expected

Code	Full	Explanation
As expected	As expected	Interaction went as participant expected
Neutral	Neutral	Interaction went not as expected but not as unexpected
Unexp.-gain	Unexpected gain	Participant did not expect the gain to be set at the level it was set at
Unexp.-cursor visibility	Unexpected cursor visibility	Participant did not expect the cursor to disappear after the button had been clicked
Unexp.-mouse weight	Unexpected mouse weight	Participant did not expect the mouse weight to be as it was

Question 5 - Influence of Interaction

Code	Full	Explanation
No influence	No influence	Participant did not subjectively believed that distance estimates were influenced by the interaction with the mouse.
Gain	Gain	Participant believed that gain may have influenced distance estimates.
Weight	Weight	Participant believed that mouse weight may have influenced distance estimates.
Influence-nonspecific	Influence-nonspecific	Participant believed that interaction with the mouse had influenced distance estimates but did not provide a clear hypothesis or mechanism
Neutral	Neutral	Participant did not provide belief nor disbelief that interaction influenced distance estimates.

Experiment 2

Question 1 - Direct Experimental Suspicions

Code	Full	Explanation
Generic	Generic	Participant made a generic statement that contained no specific hypothesis or goal.
Dontknow	Don't know	Participant expressed not knowing the experimental goal
InflPrevDist	Influence previous distance	Participant reported thinking experiment was about determining correlations between estimates
Presentation types	Presentation types	Participants suspected a difference between X and Y or exocentric and egocentric perception
distinVE	Distance in Virtual Environments	Participant reported believing that the experiment was about estimating distances in VE
Embodiment	Embodiment	Participant discussed possible influence of embodiment on estimations.
Other	Other	Other, infrequently mentioned, goals or hypotheses.

Question 2 – Motion Controller interaction as expected

Code	Full	Explanation
As expected	As expected	Participant reported interaction following expectations.
Neutral	Neutral	Participant did not have expectations or rated the interaction as neutral following expectations.
Unexp. – gain	Unexpected gain	Participant described mismatch between expected movement required and actual movement required to breach distance
Unexp. – depth	Unexpected depth	Participant did not expect depth in interaction or expected a different implementation of depth
Unexp. – controllertype	Unexpected controller type	Participant reported expecting a different controller type to be used.
Unexp. – Difficult	Unexpected difficult	Participant indicated that the interaction was difficult.
Unexp. – Weight	Unexpected weight	Participant indicated that the controller was heavy.

Question 5 - Influence of Interaction

Code	Full	Explanation
No influence	No influence	Participant did not subjectively believed that distance estimates were influenced by the interaction with the mouse.
Neutral	Neutral	Participant has no belief regarding an influence of interaction.
embodiedCue	Embodied cue	Participant expressed belief of being influenced by embodied cues (e.g. amount of arm movement or reaching)
Gain	Gain	Participant specifically referenced discrepancy between distances perceived and those traveled with the controller.
Effort	Effort	Participant indicated interaction effort may have influenced response.
Influence-nonspecific	Influence-nonspecific	Participant beliefs interaction impacted estimates without becoming specific about the mechanism for this influence and/or the consequences of the influence.
Other	Other	Participant provided one or more specific reasons that were infrequently mentioned.

Appendix D: Instruction Form Experiment 2

Bedankt voor je deelname aan dit experiment. In dit onderzoek zijn we geïnteresseerd in de rol van omgevingsimmersie bij visuele perceptie. Hierbij richten we ons specifiek op afstandsperceptie. Hieronder wordt de taak die je zo zult gaan doen in detail beschreven, lees deze beschrijving alsjeblieft aandachtig door. Nadat je dit hebt gedaan zal het Head Mounted Display worden opgezet en ontvang je een 3D bewegingscontroller. Dit kunt je zien als een soort geavanceerde Wii afstandsbediening die in 3D ruimte werkt. Hiermee voer je de taak uit.

Straks neem je plaats achter een virtuele tafel in een virtuele wereld. In deel 1 van het experiment is het de bedoeling dat je de afstand tussen twee objecten die op de virtuele tafel verschijnen inschat. Voor elke nieuwe taak wordt in de omgeving steeds eerst een rode bal gepresenteerd.

Figuur 1 De rode bal op de tafel

Om een trial te starten gebruik je de controller om het witte cursor balletje in deze rode bal te plaatsen. Er verschijnt dan een kubus op de tafel. Beweeg de cursor in de kubus. Zodra de cursor in de kubus verdwijnt zal deze groen kleuren en zal er een kort geluid worden afgespeeld.

Figuur 2 Het witte cursor balletje en de kubus

Figuur 3 Het groene balletje verdwijnt in de kubus

Druk nu met je duim op de knop bovenop het apparaat. Hierna zal de kubus verdwijnen. Onthoud de positie van de kubus en beweeg de cursor weg van de locatie van de kubus. Zodra cursor ver genoeg bewogen is zal er een tweede kubus verschijnen. Beweeg nu de cursor direct in de tweede kubus en druk op de knop. Geef dan hardop een mondelinge schatting van de afstand tussen het midden van de eerste en het midden van de tweede kubus in centimeters. Ga hierbij steeds uit van je eerste ingeving.

Z.O.Z.

Samenvattend:

1. Beweeg het witte cursor balletje in de rode bal
2. Beweeg de cursor in de kubus en druk op de knop op de controller
3. Onthoudt de locatie van de kubus en beweeg de cursor weg van deze locatie
4. Beweeg de cursor in de tweede kubus en druk op de knop
5. Geef hardop een mondelinge schatting van de afstand tussen de kubussen in centimeters

Roep nu de onderzoeker. Mocht je iets niet begrijpen dan kun je nu ook om extra uitleg vragen.

Appendix E: Fatigue Question

Vragenlijst

* Required

Uw huidige fysieke staat

Beantwoord de vraag met behulp van de onderstaande schaal:

0 - Helemaal niets
0.5 - Zeer, zeer laag
1 - Zeer laag
2 - Laag
3 - Middelmatig
4 - Enigszins hoog
5 - Hoog
6
7 - Zeer hoog
8
9
10 - Zeer, zeer hoog

In welke mate bent u op dit moment fysiek vermoeid? *

Kies een passende getal op basis van de bovenstaande schaal. Er mag gebruik worden gemaakt van getallen achter de komma en getallen boven de 10.

This is a required question

« Back | Continue »