Mind the gap:

Extending the body into 3d environments using 2d tools for interaction

Ari Kolbeinsson
Mind the gap: Extending the body into 3d environments using 2d tools for interaction

Submitted by Ari Kolbeinsson to the University of Skövde as a final year project towards the degree of B.Sc. in the School of Humanities and Informatics. The project has been supervised by Jana Rambusch.

Date

I hereby certify that all material in this final year project which is not my own work has been identified and that no work is included for which a degree has already been conferred on me.

Signature: __________________________________________
Mind the gap: Extending the body into 3d environments using 2d tools for interaction

Ari Kolbeinsson

Abstract

This thesis is a literature study on how existing research on embodied tool use may support the use of the use of the computer mouse within three dimensional environments, followed by an analysis of a typical scenario in the use of three dimensional environment. Problems with interaction in this domain are well known to designers of 3d programs but not well understood, which results in programs in which mouse controllers are used to control three dimensional objects being more difficult to learn and less efficient to use than would be possible if the interaction was better understood. The problems are often identified by their symptoms, such as the drag-threshold problem, picking problem, and the object rotation/viewpoint management problem, but this thesis will explore what the cause of those problems is, and identifies them all as a single cognitive problem which is found to be caused by a rift between the functioning of the two dimensional tool in use (the mouse and cursor) and the simulated three dimensional environment with which the cursor is interacting. Analyses are performed on a scenario, and result in a pinpointing of the problem and possible solutions to the interaction part of the problem (with design guidelines emerging), as well as finding the possibility that the cognitive roots of the problem result from an incompatibility between body-schema frames of reference for movement between the two dimensional parts of the action and the three dimensional part of the action.

Keywords: Interaction, virtual environments, tool use, embodied cognition, picking problem, drag-threshold problem, six axis, 3d mouse, Flatland, body-schema.
# Contents

1. Introduction ................................................................................................................. 1

   1.1. Background ........................................................................................................... 1

1.2. Concepts .................................................................................................................... 4

   1.2.1. The domain ....................................................................................................... 4

   1.2.2. Artefact ............................................................................................................. 4

   1.2.3. Tool .................................................................................................................... 4

   1.2.4. Affordance ........................................................................................................ 5

   1.2.5. Action possibility .............................................................................................. 6

   1.2.6. Situated cognition ............................................................................................. 7

   1.2.7. Embodied cognition ......................................................................................... 7

   1.2.8. Perception ......................................................................................................... 7

   1.2.9. Layers ............................................................................................................... 7

   1.2.10. Learning tool use ............................................................................................ 14

   1.2.11. Embodied and situated learning ..................................................................... 15

   1.2.12. Body schema ................................................................................................... 17

1.3. Aim and objectives ................................................................................................... 18

2. Methodology ................................................................................................................ 22

   2.1. Methodology of the analysis ................................................................................. 25

   2.2. The scenario defined ............................................................................................. 27

       2.2.1. Personas .......................................................................................................... 29

3. Analysis ........................................................................................................................ 31

   3.1. Analysis from a high level cognition perspective ..................................................... 31

   3.2. Analysis from an affordance viewpoint .................................................................. 33

   3.3. Analysis using perception-action coupling ............................................................ 38

   3.4. Comparison of analyses ....................................................................................... 39

4. Conclusions .................................................................................................................. 41

   4.1. Discussion .............................................................................................................. 46

   References .................................................................................................................... 50
1. Introduction

This thesis will examine how information in existing research on the extension of the human mind and body through tool use on the one hand and virtual environments on the other hand can be used together to further the understanding of how tool use extends our capabilities for performing actions into virtual environments when using a conventional computer interaction setup (mouse, keyboard and screen). The focus will then gradually be narrowed further after examining the theoretical background and introducing relevant concepts and research.

Embodied and situated cognition was identified as a central theme within a number of articles on tool use, learning tool use, and virtual environments, with the sub-fields of affordance and perception-action coupling appearing repeatedly. These identified approaches were then used to perform an analysis of a scenario which was designed to represent basic actions within a hypothetical three dimensional computer program. Personas were used to perform the analysis, allowing the researcher to go through the scenario from the viewpoint of a novice and an expert to see how well the chosen analysis methods supported the action, and some minor triangulation was done by having a novice perform the same sequence of actions in a real three dimensional program (Google Sketchup).

The results indicated that the chosen analysis methods did not fully explain how a person processes the part of the action where two dimensional movement gets translated to three dimensional movement, but did pinpoint where and why the problem occurs. This was found to be due to a break in the continuity in the action/perception due to the differing dimensionality of the offered perception and the interaction tool (mouse and cursor). Solutions are then offered for this interaction problem, which may give some guidelines to user interface (UI) designers for bypassing this interaction problem by designing the user interface in such a way that it offers interactions that match the interaction tool (the mouse and cursor). Further examination using a concept called body schema, found during the course of the literature study, suggests that the cognitive roots of the problem may be due to incompatible frames of reference between the two and three dimensional parts of the interaction.

1.1. Background

To start off with a very basic illustration: William Gaver (1991) states that “onscreen buttons seem to protrude from the screen; they afford pushing, but not moving or editing” (Gaver 1991, p.81). This may at first sight seem reasonable, especially when the interface behaves in a way similar to the tool that is used such as regular two dimensional computer interfaces with buttons that mimic the behaviour of the buttons on the mouse (press the mouse button and the on-screen button appears to get pressed). Rambusch (2010) brings up a question to a similar scenario in a three dimensional environment that involves opening a door in a computer game, and after noting that the
controller buttons (or keyboard buttons) always provide the same possible actions, pressing, that:

“..What we do know, however, is that while the player plays the game, she temporarily attaches an object (the keyboard) to her body, and thereby extends her action capabilities. In other words, the attachment of this specific object allows her to ‘reach’ into the virtual world – but it still does not explain why, or what affords the pressing of the button.”

(Rambusch 2010, p.66)

This then is illustrative of the problem that will be studied here. In the nineteen years between Gaver (1991) and Rambusch (2010) publishing their papers this question remains unanswered and remains a problem, which Rambusch (2010) informally dubs the affordance-action gap. How do virtual objects (two dimensional or three dimensional) suggest what their function is and how it should be logical to interact with them using the supplied interaction tools (keyboard and mouse for input, two dimensional monitor for output)? This implies that the cognitive process of understanding this action is not fully understood.

To break this down further, the graphically simulated virtual-world within the computer can consist of two-dimensional parts and simulated three-dimensional parts. The two-dimensional parts are often parts of the programs explicit interface (toolbars, control panels, etc.) which exists solely to assist the user in achieving a goal (Cooper, Reimann & Cronin, 2007), while any simulated three-dimensional object is often a part of the goal itself; some arrangement of the simulated three-dimensional objects may even comprise the goal itself. In some cases additional interface elements are overlaid on the simulated three-dimensional objects, with those interface elements sometimes being two dimensional in nature, and sometimes three dimensional (Cooper et al., 2007). All these elements are then displayed on a flat, vertical square (the front surface of the computer monitor) as a two dimensional grid of coloured lights commonly known as pixels. The computer mouse input device gives only two dimensional relative location information (in the form of “movement 1 unit in direction X and 2 units in direction Y”) with information neither being given about orientation nor information about absolute position (Cooper et al., 2007).

Calling the on-screen three-dimensional elements “simulated three-dimensional” is done for the simple reason that the computer works to make a two-dimensional image which can be shown on a screen but which should try to simulate three dimensions by having distance represented by scaling of objects and elevation in the visual field presented by the screen. This two dimensional representation may not elicit the same response as would the object being simulated. In fact, some virtual objects have been shown to elicit different neural responses from those elicited by the real object (Perani, Fazio, Borghese, Tettamanti, Ferrari, Decety, & Gilardi, 2001). It is unclear whether this is due to an imperfect simulation of the objects or due to some differences inherent in the simulation. Gibson (1986) discusses how movement causes the optic array (what we see) to flow outwards from the point we are travelling towards, with a magnification occurring from that point. As people are
generally moving towards a point on, or close to, the horizon (at least when moving on the ground) this point tends to be situated close to the centre of the computer screen when creating virtual environments. Performing actions in this simulated world requires extending the human body in some way into the virtual environment, and this is done through tool use. Tool use has lately been researched from both an embodied and a situated standpoint by as Susi (2005, 2006) with Lindblom (2007) pointing out that research on extending the body through tool use should focus on the act of adding the tool to the body, the “embodying” of the tool.

Designers, engineers and architects can spend large portions of their workday interacting with computer aided design (CAD) programs, which support the creation and manipulation of virtual objects. Any unnecessary inefficiency in manipulating objects in these programs can be expensive as the basis of CAD programs is manipulating simulated three dimensional objects. Many actions repeated over the course of a day with each action being less efficient than it could be leads to large amounts of time being wasted each day. A whitepaper by the Technology Assessment Group (2008) states that up to a 21% increase in efficiency can be expected by switching out the two dimensional mouse when using CAD programs and using a specialised three dimensional controller instead, after a period of learning the new interaction method. This suggests that the cost of working in a simulated three dimensional world with a two dimensional controller is quite large, with around 17% of the total time of manipulating objects, according to the Technology Assessment Group’s (2008) numbers. As three dimensional controllers are expensive, with controllers from a company called 3dConnexion ranging from USD99 to USD399 (3dConnexion, 2011) and are not likely to replace the mouse controller even in many engineering firms. The ubiquity of the computer mouse illustrates the necessity of understanding how best to support use of that interaction method in CAD programs. An understanding of how novices can learn interacting with CAD systems, as well as an understanding of how experienced users cope with a limited two dimensional controller is a key to providing guidance to the designers of future CAD systems so they can optimise interaction around the mouse controller.

Although this manifestation of the problem of moving between two dimensions and three is new, the difficulty of understanding dimensions is not a new subject and neither are ruminations on this subject. The novel Flatland was written by Edwin A. Abbot in 1884 (Abbot, 1884/1992) and is an exploration of just the complexities involved in trying to understand higher dimensions from a lower dimensions. It specifically focuses mostly on a two dimensional character, named A Square, trying to understand three dimensions and then trying to explain to a one dimensional character the concept of two dimensions. This old novel is still used by some teachers to explain these problems to mathematics students, with movie and animated versions also being used and promoted for teaching purposes (Lester, 2011). Flatland (Abbot, 1884/1992) is pertinent here as it illustrates a part of the problem being examined here; trying to interpret the relationship between the three dimensional action which is the objective of the task, and the two dimensional tool movements required to achieve that action.
The preceding segment illustrates the complexity of reaching into the virtual world. It also illustrates some differences between the presentation of virtual and real objects, and introduces tools as an extension to the body. The following chapter will introduce further concepts which are required to understand the domain and the interaction problem being examined.

1.2. Concepts

Some concepts important to understanding the thesis will be defined here, collected in one place for easy referral during the reading of the thesis. These particular concepts are collected here because the following chapters rely heavily on the reader having at least a passing familiarity with these concepts and their history. The concepts action possibility and affordance are also connected to such a degree that introducing them in quick succession is logical.

1.2.1. The domain

The domain being examined is a regular computer running a program showing three dimensional objects on a screen and using a mouse to provide input and interaction with those objects. That is, all interaction is done through the typical tools available to most computer users.

1.2.2. Artefact

An artefact is an “object made by a human being, typically an item of cultural or historical interest” (Concise Oxford English dictionary, 2008). This is a very broad concept, and is defined here for completeness as it is often used together with the concept “tool”. Susi (2006) points out that multiple definitions of the term artefact exist, and these can cause some confusion when using the term as there can be nuanced differences separating definitions. Susi (2006) does find a common meaning, which is in line with the definition that was listed here above.

1.2.3. Tool

A tool is defined as: “1. A device or implement used to carry out a particular function. 2. A thing used to help perform a job.” (Concise Oxford English dictionary, 2008). Susi (2006) points out that two principal definitions can be found for the concept. On the one hand tool refers to manually operated devices, on the other hand a tool can be almost anything since definitions of a tool include ‘anything used in any pursuit’ (Susi 2006, p.28).

Tool use in general has been shown to be learned, all the way down to the most basic tool use such as dragging an object towards us by dragging on the cloth the object stands on (Schlesinger & Langer, 1999) and Merleau-Ponty (1945/2005) notes that learning tool use depends on the situation in which he sees the tool being used.
Experiments done by Schlesinger and Langer (1999) were done on infants with infants at 8 months age only performing directed actions with connected tools, but 12 month old infants had sufficiently trained their sensorimotor- and cognitive systems to such an extent that they saw the action possibility inherent in the the more complex tool scenario. This suggests that even the basic understanding of how to manipulate objects is learned, in the sense that all sensorimotor development results from practice, and suggests that action possibilities, and the perception thereof rely on learning for even the most basic of tasks involving tool use. Schlesinger and Langer’s (1999) work also supports recent work by Adolph, Eppler and Gibson (1993) on infants’ understanding of their own capability for performing climbing and descending actions.

1.2.4. Affordance

Affordance is a concept formed by James J. Gibson in the nineteen seventies and made popular in ecological psychology and psychology by his 1979 book *The ecological approach to visual perception*. Affordance refers to the attributes of an object or the environment that let us interact with those objects or environments (Gibson, 1986) or as Gibson himself puts it “The affordances of the environment are what it offers the animal, what it provides or furnishes, either for good or for ill” (all italics from the original text, Gibson, 1979, pg. 127). Gibson (1986) contends that these affordances exist in the objects themselves and that they are directly perceived by the observer without intermediate cognitive steps. Gibson (1986) mentions that the body is involved in perception but gives only a partial explanation of this. Affordances thus seem to be an attribute, information about an attribute, and the transmission of that information, according to Gibson’s (1986) own text.

Donald Norman (1986) then popularised the term affordance, but with a different meaning: “the real and perceived attributes of a thing, primarily those fundamental properties that determine just how the thing could possibly be used” (Norman 1986, p.9). Norman (1986) did not agree with Gibson’s (1986) view on the direct perception of affordances, opting instead to combine a concept of *mental models*, which refer to our already gathered knowledge on how the world works, with affordances. This is a fundamentally different approach to Gibson’s (1986) original concept of direct perception and the two are not easily reconcilable.

The applicability of the whole concept *affordance* to virtual objects has also been questioned (Norman, 1999) but researchers such as Rambusch (2010) and Gaver (1991) see the term as useful even for virtual environments and use it in research on virtual objects.

Martin Oliver (2005) even argues that the multiple definitions of affordance and the rampant misunderstanding and misuse of the concept render it unfit for use, especially as it can create an illusion of a scientific consensus where none exists both within fields such as human computer interaction (HCI) and to those in related fields, as well as non-experts (the layman).
The term *affordance* will be used when discussing an object’s action possibilities coupled with a model for making that information useful for action. As the term affordance can refer to more than one theory/model every effort will be made to ensure that it is clear which version is being discussed at each time.

1.2.5. Action possibility

An action possibility is, quite simply, a possibility for action offered by an object (McGrenere & Ho, 2000). This is in essence the actual property of the object which is often described by the term *affordance*, yet the term *action possibility* does not itself belong to any specific framework or theory for explaining how the action possibility gets perceived, interpreted or acted upon by an actor.

The term *action possibility* is also sometimes used to explain the term *affordance*, with a popular definition residing on Wikipedia: “The original definition described all action possibilities that are physically possible” (Wikipedia, 2011). The Wikipedia article is interesting in this context as it showcases the popular understanding of affordance. Turning to peer reviewed material we find McGrenere and Ho (2000) stating that: “Gibson intended an affordance to mean an action possibility available in the environment to an individual, independent of the individual’s ability to perceive this possibility” (McGrenere & Ho 2000, p.1). Dohn (2009) has criticised how McGrenere and Ho (2000) use the term “action possibility” due to McGrenere and Ho (2000) failing to specify how the individual’s action capabilities work. The term *action possibility* can still be seen as useful as it is not surrounded by controversy and confusion to the same degree as the term affordance as well as having no specific theory of how the information is transmitted or processed. This makes the term action possibility modular in nature and easier to incorporate into new theories of perception than affordance which comes with a history of multiple interpretations. One of the problems with affordance is that the term is used to describe many things, even in Gibson’s (1986) original text affordance is used to describe an object’s attribute (that an object is graspable) as well as to further describe that attribute information (how it is graspable and by whom), and the transmission of that attribute information (how that information gets to a person and can be acted upon).

Keeping action possibility as a more limited and purely descriptive term would make it possible to work in a more modular fashion, figuring out one part of the puzzle at a time. This approach would also allow for changes in each part of any possibly emerging theory without being forced to throw out the baby with the bathwater. That is one problem with affordance, that because the term encompasses almost all aspects of Gibson’s theory as well as his model of ecological perception it becomes hard to revise any one of these elements when new research becomes available (Oliver, 2005).

The concept *action possibility* is thus used in this thesis when the goal is only to discuss what action(s) can be performed with/on the object, but not to
discuss a complete model or theory of how information about that possible action passes between the object and the actor.

1.2.6. Situated cognition
Situated cognition is a theoretical approach that views the role of interaction between a cognising human agent and his environment, both with the physical and the social environments, as being central to human cognition (Susi, 2006). Situated cognition is often combined with embodied cognition, both of which will be examined further throughout this thesis.

1.2.7. Embodied cognition
Embodied cognition refers to the mind and the body being inseparable entities with cognition being a property of the mind coupled with the body’s sensorimotor system (Lindblom, 2007). An important note here is that according to the embodied cognition viewpoint all cognition is embodied.

Situated and embodied cognitive viewpoints are commonly combined to form a view on cognition wherein the interplay between environment, mind and the body (sensorimotor system) is a crucial and inseparable part of cognition and has been split into a high-level component which consists of socio-cultural factors and prior internalised learning, context which consists of the environment, the here and now, and finally low level which consists of the mechanics of the cognition, the hardware, which in this case is the sensing apparatus (eyes, ears, touch, proprioception), the cognitive apparatus (the brain/mind) and the action apparatus (the hands) (Susi, 2006; Lindblom & Susi, in press). Emphasis will be put on visual perception in this thesis, as the focus is on interaction with a virtual environment with no simulation of other types of feedback.

1.2.8. Perception
Perception has a general definition which revolves around becoming aware of something through the use of the senses (Concise Oxford English dictionary, 2008). A more specific definition is required for this thesis:

“Perception
• Psychology & Zoology - the neurophysiological processes, including memory, by which an organism becomes aware of and interprets external stimuli.” (Concise Oxford English dictionary, 2008)

Perception will be further discussed, explored and defined throughout the thesis.

1.2.9. Layers
As this thesis focuses on interaction with three dimensional virtual environment it becomes necessary to find a way of describing and discussing that interaction. The layers concept attempts to create a visual and easily
understandable way of disassembling the tool being used into components. This is needed for the analysis of complete actions.

Any object created in a typical computer program for dealing with three dimensional objects is either a part of the regular two dimensional computer program interface, or part of the three dimensional “virtual world”.

Any object in the real world has certain attributes, some of which are visual in nature and others which may only be discovered using other senses, but the objects created in the computer (virtual objects) can only give off visual information about their attributes (at least through the screen, keyboard and mouse combination. Objects must thus give all information for interaction, as well as all feedback from interaction, through only these limited channels. Further adding to the limitations is the fact that the screen is a two dimensional visualisation device (traditionally aligned in the vertical plane) and the mouse is a two dimensional pointing device aligned in the horizontal plane. The mouse is two dimensional because it only gives information about its location on the x and y axes, but does not give information about its orientation (where it points) or any motion in the z axis (depth, into the screen).

Virtual objects lack a physical presence, they lack corporeality, but they can share some, or all visual attributes of a real object. This lack of corporeality still means that any interactions we have with virtual objects are tool based and can therefore be quite different from interaction with real objects unless highly specialised and/or task specific interaction technologies such as haptic interfaces or custom controllers are used. These specialised interaction
technologies can be specific only to that task, expensive, and can require whole rooms (a projected virtual environment with a bicycle used for control, for example). A range of devices also exists purely for interacting with three dimensional programs in a desktop computing setting. A small selection of these devices is shown in Figure 1. These devices include highly specialised devices such as the 3dConnexion SpaceNavigator, the Labtech SpaceBall 4000, and the Axsotic 3D-Spheric-Mouse as well as the more conventional Sandio 3d Game O2 mouse, which merely adds two small joysticks to a regular mouse for more axes of control. These devices are not commonly used, as they are not available everywhere, expensive, can be uncomfortable for extended use, and can be of limited use in regular computing. This thesis focuses on some problems inherent in the most common of controllers, the computer mouse, which is commonly used for controlling all types of applications ranging from computer games to office applications to computer aided design (CAD) software, while recognising that specialised controllers likely change the nature of the problem being examined or eliminate it entirely. The fact remains that the mouse controller is the de-facto standard controller, and is used by most computer users in all tasks, and in design offices and school labs where future CAD users learn.

An important question is: how can we translate the perceptible possibilities for performing actions on virtual objects in the virtual domain to the actions we need to perform in the physical world given the limitations imposed by the tool used for interaction. An answer to that question would support the design of three dimensional programs such as Computer Aided Design (CAD) packages as well as add to the understanding of human-computer interactions.
interaction (HCI). Figure 2 is an illustration of this problem, and shows these different parts of the required interactions as *layers*. This concept was created by the author just to illustrate this problem. The layer concept was created purely as a method of visualising the multiple areas where action and perception have to occur for goal directed actions to be possible within the virtual environment, and does not imply a model or theory to explain how those actions are performed.

Layer 1: The physical world. Basic input/output such as the screen, the keyboard and the mouse. See Figure 2, pg. 9.

Layer 2: Application controls such as buttons and menus for interfacing with the program itself. This layer is two dimensional even though shading may be done on individual interface elements to suggest some three dimensionality, particularly when an action is
performed through a button click. The mouse cursor can be seen to exist in this layer, as the cursor is two dimensional in nature and moves only in the plane of the two dimensional interface. See Figure 2, pg. 9.

Layer 3: Controls for manipulating simulated three dimensional objects. These controls are three dimensional in nature, and show a direct link to the object to be manipulated. See Figure 2, pg. 9.

Layer 4: The three dimensional virtual objects themselves. All the other layers are made for the purpose of interacting with this layer. See Figure 2, pg. 9.

As layers three and four provide different visualisations of action possibilities and different constraints on actions they will be treated as separate layers in this thesis, even though they are co-located in three dimensional virtual space. Layers three and four follow largely the same rules, existing as they do in the three dimensional virtual environment. Even so, one important difference separates layers three and four, which is that layer three is a tool layer while layer four consists of the environment and objects, the manipulation of which is the goal of the system.

Another element to visualise is Rambusch’s (2010) question, introduced on page 1, which has to do with the problem of how the control unit allows the player/user to reach into the virtual world. Figures 3 and 4 show how a user might use the mouse cursor to select elements within the virtual environment, but in Figure 3 this is shown from the user’s perspective while Figure 4 illustrates the same action in terms of the aforementioned layers (with layer 3 left out as it is not needed in this illustration). The mouse cursor acts as a two dimensional element and uses no cues to suggest that it moves “into” the environment, yet it allows the user to interact with the virtual object layer.
(layer 4) as well as any specific layer 3 controls that might appear as needed. The cursor can thus be said to be constrained in motion to the two dimensional layer 2. This gap between layer 2 on the one hand, and layers 3 and 4 on the other is the place where traditional tool use theories may have problems seeing as how the tool and the object to be manipulated exist in different domains, the tool (the mouse and the cursor) in a two dimensional space but the object in a three dimensional space. There is nothing to suggest how the tool (the cursor) connects with any object within the three dimensional realm, as the cursor does not obey the rules of the three dimensional realm such as being shown to get smaller as it reaches “into” the screen, nor is the user given any way of actually moving the cursor in more than two dimensions. This contrasts starkly with tool use in the real world where the tool follows the same rules as other artefacts in the scenario in which the tool is being used. Figure 5 illustrates this in a simple manner, showing a ruler being used as a tool to extend a person’s reach to operate a light switch. The light switch gives cues to its distance by seeming smaller or larger based on the user’s distance and the ruler shows the same characteristics, with any point of the ruler which is closer to the user seeming larger than an equally large point of the ruler which is more distant. The computer cursor behaves in a way more like that shown in Figure 6. Figure 6 shows a ruler being held and manipulated only along the X and Y axes (left-right and up-down) and overlaid over the light switch in much the same way as the cursor in Figure 3 (pg 10), and Figure 7 (pg 11.) shows the same angled view of this scenario as is shown in Figure 4 (pg. 10). This is simply not how tools work in the real world. It is already apparent that tools do work in this way in virtual environments, but not readily apparent how this gap between layer 2 and layer 4 is bridged. Note that this applies to any scenario where a two dimensional interaction technique is used for interacting with a three dimensional virtual world; touch-screen based interaction requires the same bridging, as the interactive surface is two dimensional but the virtual environment stretches into the screen. The cursor is not directly linked to the
Figure 9: The cursor is being moved at an angle of approximately 20° from vertical. The selected cube is moving through the other cube.

Figure 10: The cursor was moved at an angle of approximately 25° from vertical. The selected cube has moved horizontally towards the viewpoint.

Figure 11: The cursor was moved at an angle of approximately 25° from vertical. The selected cube has moved horizontally towards the viewpoint.
object either. Instead, a temporary link is created through button presses. As there is nothing to directly indicate how this link between the physical mouse button press and the virtual button press it is necessary to use some explanatory model such as the user creating (or having a pre-formed) mental model of how the action will transfer or a more direct perception-action loop based model which views all parts of the action as being exploratory in nature with any small action giving feedback through perception, thus making further action possible. This is illustrated in Figures 8 through 11 using the freely available program Google Sketchup. Figure 8 shows a starting point at which the user has selected the rear cube (shown in the program by bold, blue lines and dots on the surfaces) and has the “move” tool selected (a cross of arrows). The mouse button is then clicked, and the cursor dragged at a two dimensional angle of around 20° from vertical, see Figure 9. This movement causes the cube to move “towards” the user and scale bigger, while the cursor move tool moves only at the aforementioned 20°angle and does not change size. Figure 9 also shows the cube moving through the cube which was partially occluding the selected cube. The angle of the cursor movement was not perfectly stable throughout the action and in Figure 10 the movement is completed as a 25° angled movement of the cursor (from vertical). Figure 11 then illustrates what happens if the mouse movement goes nearer to vertical. That causes the object movement to suddenly change to a movement along the vertical axis. This means that if the user wants to move the object left or right at the same time as towards or from himself then the object can suddenly switch movement axes. This particular behaviour is particular to the interaction method used in this program, Google Sketchup, which illustrates one attempt at solving the problem of moving three dimensional objects using a two dimensional interaction tool. This approach has the advantage of making more axes of movement available to the mouse user, but the downside is a certain lack of predictability as well as quite high demands on motor skills as the thresholds for switching axes can feel quite narrow.

It is worth pointing out that the problem being discussed is known and referred to as separate problems of design and action (the picking problem, the drag threshold problem, and the object rotation/viewpoint movement problem), but is treated as a void in books on design (Cooper et al., 2007 and Benyon, Turner & Turner, 2005) with no mention of a solution or any research into why the problem occurs.

1.2.10. Learning tool use

As this thesis will attempt to understand and explain how a person learns complex virtual tool use in the form of manipulating a simulated three dimensional object in a virtual environment by using a two dimensional mouse controller it is necessary to have some understanding of how humans learn tool use to begin with. To understand the analyses that are examined in chapter 3 it is crucial to have some understanding of how humans piece together their understanding of how the world works in relation to their own bodies and how humans learn how to relate the environment and objects to their own bodies and capabilities.
How children go about learning what actions are made possible by an object or environment has been extensively studied in the real world, and is a relevant examination of how tools are used and learned without prior experience. Learning tool use is based on gaining an understanding of objects relations to each other and how they can affect each other. As any interaction with a computer requires the use of a specialised tool of some sort, in this case we focus on the computer mouse, how tool use is learned bears some examination. Furthermore, an important part of the question as to how we can use a two dimensional controller in a three dimensional environment has to do with how we learn to perform the task.

Adolph et al. (1993) found that toddlers would treat a sloped downward surface (20° slope, and 30° slope) with caution and shift to a safe sliding position before advancing, while crawling infants would generally plunge headfirst straight down. The toddlers would furthermore try out various positions for stability before going down. When faced with an upwards slope all the infants would try out ascent techniques, a learning by trial and an exploration of the environment. Adolph et al., (1993) are careful in their interpretation but conclude that this difference in behaviour is likely due to the infants not yet having learned how to interpret the environment’s affordances in relation to their capabilities for interacting with the environment and performing actions, whereas the toddlers have already gone through a period where this type of perception-action couplings are explored. This shows the importance of experience and suggests that some form of mental models are required, wherein a person can attempt to fit a new situation into previous experience which may even come from other domains.

Schlesinger and Langer (1999) later found that infants learn to perform actions with tools before they perceive causal links in watched actions. These findings are then strengthened by research done by Sommerville, Woodward and Needham (2004) which focused on infants who got to perform a task either before watching a third party perform the action or after watching the task being performed by a third party. Sommerville et al. (2004) found that the children who got to perform the task before watching it performed were significantly more attentive when watching, suggesting that performing an action helps the infant understand another actor’s goal directed actions. This can be seen as being mostly in line with Rizzolatti and Craighero’s (2004) work on mirror-neurons, which revolves around a system of neurons which facilitate the imitation of observed actions, yet Rizzolatti and Craighero (2004) worked with actions with which test subjects, whether human or simian, were already familiar. Sommerville et al.’s (2004) results suggest that performing the action may have an effect on any subsequent perception of the action. As this shows the connection between doing and seeing it can be seen to support a body based view on perception, which will be further explored when body schema will be introduced in chapter 1.2.12 on page 17.

1.2.11. Embodied and situated learning

Andy Clark’s (1997) explanation of a loop of perception and action comes in with an interesting view on exactly this “learning to perceive through doing” which has just been described. Clark (1997) basically posits that simple rules
for action are learned early on, and that that actions built upon these will always rely on the embodied and situated nature of the learning process. Understanding of possible action by the body and mind in the environment is thus a function of prior actions performed by the body and mind in functionally compatible environments. Clark (1997) goes one step further and suggests that even more complex and seemingly non-body based cognition (possibly even seemingly logical and propositional) relies heavily on body based metaphors related to motion, and learned through performing actions.

Gibson (1986) rejected both the traditional stimulus-response approach to perception as well as the (then emerging) information processing view of perception, opting on a situated view of perception which views the actor’s perception as being an integral part of the surrounding environment. Although Gibson incorporated a situated view of perception his rejection of the information processing view and indeed, as Clark (1997) points out, Gibson’s rejection of any inner state based processing, causes problems because even action-perception loop based embodied approaches require basic inner state based processing of environment information (Clark, 1997). Inner state based processing refers to some sort of internal representation of an object coupled with internal manipulation of the representation.

Clark (1997) gives Gibson the benefit of the doubt and treats his rejection of inner state based processing as a possible slip in Gibson’s wording, rather than his thought, and points out that Gibson’s words are generally interpreted in a looser way as opposing only a fully representative notion of internal processing, as can be seen in Gaver’s (1991) work on perceiving affordances in computer programs. Gaver’s (1991) focus was on how affordances could be used to explain complex actions by viewing nested affordances, or affordances that may reveal themselves only when a previous affordance is used. Many actions can be explained by this, an apple may afford biting, but it is only after taking a bite that chewing becomes possible. Gaver (1991) also commented on how even complex actions could thus be explained as a series of simpler actions, obviating the need for cognition intensive processing as an explanation in many cases.

Perception research is in some ways a very well established field, as it has been covered by philosophers since the ancient Greeks, with Aristotle separating perception from reason and describing perception a way of receiving information about objects without having to take in the matter of the objects themselves (Internet Encyclopedia of Philosophy, 2005), but in other ways the field is less developed. This can be seen by the disagreement and confusion over basic concepts such as affordance which is apparent from the differing interpretations by Gibson (1986) Norman (1986), Gaver (1991), Dohn (2006) and others researchers within perceptual psychology or applied perception. Due to this weakness of definition it becomes important, in addition to reviewing recent research into perception, to go back to some of the older work on perception. This entails looking into the philosophy of perception and examining it with more recent research as a guiding light. As that field is extremely large and complex this exploration will be limited to some approaches which are known to work together with the embodied and
situated views of cognitive science and tool use, seeing as how this approach has been gaining support within the cognitive science community in recent years.

1.2.12. Body schema

Body schema refers to the view that our understanding of objects and motion are framed in relation to our sense and understanding of our own body, which is a view that is based on the work of the philosopher Maurice Merleau-Ponty (1945/2005). To get an understanding of what body schema is, it becomes necessary to go in as short a form as possible through the philosophical underpinnings behind body schema.

The philosopher and psychotherapist Eugene Gendlin (1992) argues that just being is an interaction with the environment, and that no perception is required for interaction. Gendlin (1992) furthermore argues that any definition of perception presupposes a perceiver, making the existence of the perceiver the first step towards interaction. A flower interacts with its environment without perception, according to Gendlin (1992), but it can be said that even natural forces cause interaction, such as in the case of rocks pushing each other due to gravity, although that interaction has no intent behind it. Interaction simply means “reciprocal action or influence” (Concise Oxford English dictionary, 2002) and is used for describing interactions which do not require any agent of intent within fields such as statistics (Shaughnessy, Zechmeister & Zechmeister, 2008) and physics, where the term describes how matter, fields or subatomic elements affect each other (Concise Oxford English dictionary, 2002). Gendlin’s (1992) view thus highlights the importance of the perceiver, and how the perceiver changes the nature of the interaction from being something that merely happens to something that is done; an action in response to a perception instead of only a chain of actions as in the case of rocks pushing each other. This lays down a dividing line between the active and non-active interaction, on which sits the perceiver and his frame of reference for perception. That active interaction is made possible by perception, and it is that dividing line which provides the concept of the body schema as the reference point to which perceptions can be anchored (Gendlin, 1992). Gendlin (1992) and Merleau-Ponty (1945/2005) argue that this frame of reference for perception is the perceivers body, and that this frame of reference is a part of the perception itself. This means that all perception, whether visual, tactile, or olfactory, is linked to how we think of our own body.

Both Gendlin (1992) and Merleau-Ponty (1945/2005) discuss movement as an integral part of perception with Merleau-Ponty (1945/2005) pointing out quite strongly how all perception is body-schema based in some way by using the problems apraxic patients (patients with problems emulating actions or performing actions on command, but with a full understanding of what they are trying to do or being asked to do) face when trying to emulate another’s actions as an example.

“..the normal subject has his body not only as a system of present positions, but besides, and thereby, as an open system of an infinite
number of equivalent positions directed to other ends. What we have called the body image is precisely this system of equivalents, this immediately given invariant whereby the different motor tasks are instantaneously transferable. It follows that it is not only an experience of my body, but an experience of my body-in-the-world, and that this is what gives a motor meaning to verbal orders. The function destroyed in apraxic disturbances is therefore a motor one” (Merleau-Ponty 1945/2005, p.163).

Merleau-Ponty (1945/2005) states that the environment, consciousness, and the body are all intertwined in such a way as to be inseparable, which goes against the Cartesian view of consciousness and body being separate but connected entities. This fits rather nicely in with the modern view on embodied and situated cognition and could thus be considered a useful point for philosophically grounding the embodied and situated viewpoints.

“When I hear a melody, each of its moments must be related to its successor, otherwise there would be no melody.” (Merleau-Ponty 1945/2005, p.474)

1.3. Aim and objectives

The purpose of this thesis is to explore how interaction with three dimensional virtual objects through two dimensional tool use can be explained from a situated and embodied standpoint and possibly comparing this interaction with non-virtual interactions to see whether action possibilities of virtual objects and real objects share similarities. This will be done through a literature study examining earlier research which covers the various aspects of the tasks involved. This includes studies on child development in regards to how tool use develops, studies on tool use in general, the use of artefacts and tools for cooperation and communication, motor control, perception within virtual environments, affordances in computer programs and more. These diverse studies are required as they provide a foundation from which to explore the specific tool use which is required for extending a two dimensional mouse controller into a simulated three dimensional world.

After reading the background information presented in the previous chapter it becomes easier to see the shape of the gap that needs bridging between two dimensional tool use and three dimensional virtual environments. Concepts which emerge on both sides of the gap between layers two (two dimensional) and four (three dimensional) are perception and action. Research into tool use must, at its core, always explain how the perception leads to action, or how action influences or leads to perception. Conversely, the studies presented on virtual environments, as well as some studies on motor skill development, commonly use the term affordance in some manner to explain possible interactions, and as affordance is a term coined to explain visual perception it is possible to see perception as being a large common factor joining the different areas of research.

A question arising from this is how tools can extend cognitive- and action capabilities of the physical body into a virtual environment. The ultimate aim
of that question is then to find a way of discovering, describing, and verifying a mechanism or theory which supports tool use in virtual environments, but that requires some intermediary steps as this is a very large question. One such smaller intermediary step centres around what similarities can be found in existing research on tool use on the one hand, and virtual environments and action possibilities/affordances on the other hand which support a bridge between the two areas. This narrower area will be the focus of the remainder of the thesis. Extrapolating a possible experimental approach from this earlier literature by finding a way to adapt existing research from one domain, such as development of tool use research, and apply it to the other domain, virtual environments, is a secondary goal. A major goal of any possible bridge identified would be to support the design of computer interfaces for working with three dimensional objects through the conventional mouse/keyboard/screen interface.

The gap that is addressed by this thesis is that no research has been found which explains how the perception of possible actions transcend the gap between the three dimensional virtual world and the tool (which is two dimensional in both the real and virtual worlds). This perception of possible actions must then loop back into the virtual worlds as actions. How does such a conceptually complex loop of perception and action end up being so easy to perform? Looking back to Figures 8 to 11 it is possible to see both how simple the execution of the actual task is, and also how easily the task can fail. The reader is invited to experiment with Google Sketchup, which is available for free through Google, or a CAD program of choice, keeping in mind the disconnect between the two dimensional movement of the cursor and the mouse and the three dimensional nature of the interaction.

Existing research within human-computer interaction, developmental psychology, and cognitive science seems to focus on either tool use in the real world, including the nesting of tools, learning basic tool use and more, or focus on what happens within the virtual realm. This latter class of research often uses the affordance concept as an explanation or a “miracle concept” for explaining all possible actions within the system, without having shown how perception or tool use transcend the gap between real and virtual. Gaver’s (1991) paper on technology affordances can be taken as an example of this, as Gaver makes statements such as “onscreen buttons seem to protrude from the screen; they afford pushing..” (Gaver 1991, p.81). That is done through techniques such as shading. What needs explanation is how using a two dimensional tool in the real world links into the three dimensional virtual worlds so that the actions within the system can preferably be explained using terms which have been developed for explaining general tool use. This lack of understanding of the translation of action and perception between these two commonly used elements within computer programs has led to the various well known three dimensional programs being unnecessarily hard to use. A whitepaper by The Technology Assessment Group (2008) shows that when this factor of a gap between action and perception is at least partially removed through the use of a specialised three dimensional controller such as the Connexion Space Navigator, work efficiency in CAD applications goes up by 21% based on self reporting from 190 users of a three dimensional controller. Note that this whitepaper may be funded by Connexion, and is
available on the Connexion website. García, Quirós, Santos and Peñín (2007) also illustrate the difficulties of learning how to use CAD programs and show a specialised CAD program for learning called GIcad which has fewer functions than a full fledged CAD program and a simplified interface. This program still requires a large amount of knowledge from the user if he is to successfully complete the tasks he wants to perform, and if the user has a time limit on performing the action then that requirement for knowledge is increased (García et al., 2007). The fact that this simplification of the program interface does not make the task easily doable by a novice suggests that the actual interaction makes for difficulties in a task that is in its essence a digital simulacrum of sculpting from clay and carving from foam and joining these objects together. Hamade, Artail, and Jaber (2007) also showed that separate the skills required for working with the relationships between on-screen objects, and for obtaining the declarative knowledge required need to be developed independently for carrying out commands specific to each CAD program. Learning this declarative knowledge has been shown to be quite difficult, especially to begin with (Hamade et al., 2007), which is a major cause of concern for those new to CAD programs García et al. 2007). Hamade et al. (2007) also found that the complexities of learning CAD programs are largely cognitive in nature, by which they mean that CAD learning requires complex mental functions, rather than just learning physical actions through repetition. This is not a new discovery as Buxton showed in 1986 (in Norman, and Draper, 1986) that interaction with different types of systems has different requirements for interaction and that CAD programs in particular have special needs for three dimensional control both for movement and for rotation, resulting in a total of six axes which require manipulation. Buxton (in Norman, and Draper, 1986) suggests that although specialised controllers which support the interaction which is required are preferable, realistically it must be remembered that computers are used for a variety of tasks, and having specialised tools for all tasks which might benefit from such controls is problematic due to space concerns, financial concerns, and simply finding and setting up the required controller for each task may not be time effective. Specialised controllers thus seem likely to remain used only by a small subset of users who use that computer primarily for a specific task, such as a computer used mainly for CAD work at an engineering firm. Other users will have to use standard controllers and will need the program to provide as much support for this limited interaction method as possible. For the program to provide support, the designers of the program must have guidelines that help them get around the problems of interaction through these standard controllers, and for those guidelines to exist it becomes necessary to understand how a user’s cognition is, or is not, supported when using standard two dimensional controllers to interact with a simulated three dimensional world.

One of the complexities of interacting with CAD programs may in fact partly be due to a trial-and-error approach being used for designing the translation from three dimensional parts of their interfaces to the two dimensional controller, due to a lack of specific guidelines. While well understood theories and guidelines exist for designing the plain two dimensional part of the interface such as menus, toolbars, dialogs and ribbons (Cooper et al., 2007), it
is interesting to note that Cooper et al. (2007) spend five pages out of 563 on the problem of three dimensional interfaces, and highlight three symptoms of the core problem explored in this thesis as being the primary problems of interacting with three dimensional object. Cooper et al. (2007) split the problem into three parts, *drag thresholds* (translating a 2d dragging motion into three dimensions), *the picking problem* (picking objects at various depths in the scene, and the *object rotation / viewpoint movement problem*. Cooper et al. (2007) spend approximately two pages on these problems, and give only some general notes as to how some program vendors have solved the problem, but not giving any guidelines as to what solutions would work, why solutions might work, or what principles may lie behind those solutions. This contrasts starkly with the rest of their book which generally gives clear instructions on how problems can be solved and why those solutions support the user (Cooper et al., 2007). Benyon et al., (2005) ignore these issues entirely in their 700 page book on designing interactive systems, although virtual worlds and special controllers are mentioned.

These books are mentioned because they are well known and because of their respective strengths. Cooper et al. (2007) is well known within the user interface design community and is widely used as a practical reference for design, and Benyon et al. (2005) is a detailed and theoretical approach to evaluation and design. The books can be seen as somewhat indicative of two popular approaches to interface design and evaluation, and both books are widely used. Between them, the two books display an almost complete lack of assistance for designers when working with three dimensional objects using the mouse controller. This makes designing this type of application unnecessarily hard and possibly somewhat hit-and-miss which can lead to inefficiency for the users using those applications, as was shown by the Technology Assessment Group (2008), as the available guidelines for design lack a theoretical grounding. This lack of a theoretical grounding is clearly illustrated by the lack of explanations given as to why the particular solution mentioned by Cooper et al., (2007) in the program Google Sketchup may be a good solution.
2. Methodology

This thesis is a literature study and examines whether the many available research papers on virtual environments on the one hand and tool use (including the learning of tool use) can be linked together to support the notion of tool use in virtual environments. This is done through backtracking to the earliest thinkers who supported the notion of embodiment to find a solid ground from which to build, then comparing tool use, perception, and learning research with research on action and perception in virtual environments. No experiments are planned, although figuring out a way of adapting either experiments around tool use for testing virtual environments, or vice-versa, would be a positive endpoint. The examined literature will then be used as a theoretical basis for and analysis of a scenario featuring the problem area. The goal of that analysis is to further pinpoint the location and nature of the problem being examined, and to find how existing theories may be used to explain the problem, either fully or partially. The scenario involves mentally going through a short and typical task within a three dimensional CAD type interface, using defined, hypothetical users as virtual (within the mind of the analyst) test subjects. Real CAD programs are used for reference to make this mental walkthrough, or role-playing, as realistic as possible, and the characters created to represent users were designed to represent two different skill levels. These techniques are commonly used together within the field of interface design and are called persona-based scenarios (Cooper et al., 2007) and are used because they allow the designer to explore the thoughts within the user’s head.

The reasoning for this methodology is simple; the gap in understanding how the perception of action possibilities occurs between the real and virtual domains means that there is a lack of previous work which can directly be referred which in turn means that any work done to examine this gap must be done based on work in adjacent fields. The problems in program design are known, but no path to solving those problems can be seen. A path must be found that can give designers and researchers pointers on which heading to take instead of having to wander around with no map, as is the case today. The amount and complexity of material that must be reviewed is such that experiments are not feasible given the timeframe. Using a persona-based scenario in this context is done to access the more complex internal dialogue of the user, and his non-verbal dialogue with the system; his back-and-forth when trying out different options, his understanding of each perception and the action that caused that perception (Cooper et al., 2007). This is not possible through conventional methods, as even interview- or questionnaire based de-briefings of non-expert subjects cannot give access to that internal monologue as it happens, but rather gives a summary of the subjects experience after the fact and is sensitive to factors such as the subjects emotional state at the time (Patton, 2002). An observation using subjects performing a think-aloud protocol might be possible, but requires a suitable testing system, and it is unlikely that this more complex and work intensive test setup would offer any information than a persona-based scenario performed by a single expert.
As this is a literature study it is crucial that prior research within the field is searched for anything which may shed light on the problem of translating action-perception between seemingly incompatible domains. This approach has primarily involved searching through available on-line databases that store academic articles, as well as examining the reference section of articles within the fields of computer science (virtual environments), cognitive science (embodied tool use), developmental psychology (learning tool use, training perception) and perceptual psychology (affordance, perception and motion). A partial list of search systems and search terms used follows:

**Search systems used:**
- Google scholar
- Academic Search Elite
- ACM Digital Library
- ScienceDirect
- SpringerLink
- My own collection of books and articles.

**Search terms used included:**
- Affordance, embodied, situated, virtual, tool use, two dimensional,
- virtual environment, mouse and keyboard, reach, extend the body, two dimensional tool use, three dimensional environment, picking problem, drag threshold, CAD, AutoCAD, Pro Engineer, Sketchup, 3d, three dimensional, interaction, usability, analysis, efficiency, as well as multiple combinations and variations on those search terms and any search terms that the author has thought of but not recorded.

These terms were used in various combinations, including search modifiers such as quotation marks and Boolean operators (AND, OR, NOT) as applicable and as supported by the various search engines. A large number of articles were read and deemed as not contributing anything useful to the area being examined in this thesis. Most articles found on human-computer interaction examining interaction with virtual worlds used custom interfaces such as bicycles or treadmills to interact with either head mounted stereoscopic displays, or projected virtual environments which project images all around the user. These articles use a domain specific interaction technique which means that the gap between layers 2 and 4 (see Figure 2) may not exist, at least not in the sense it does here. The tool matches the world which it lets the user reach into, although Rambusch (2010) demonstrated through the use of a bicycle controller that even this can easily break down if the tool does not display the expected behaviour in all respects. These domain specific controllers are an interesting study in of themselves, but the majority of computers in the world function with a relatively standard setup of a screen, a mouse, and a keyboard. This is the control scheme available to most users, and used in most cases when interacting with three dimensional environments on desktop computers. Note that even systems such as touchscreen devices do not change the issue at all, but just drop the mouse and cursor from the action. Even though a finger is used, it is still only interacting with the screen in two dimensions, with the touch sensitive element of the screen bridging the gap between the outside of the screen and the virtual world contained within. The gap between the virtual environment and the tool can thus possibly be traversed by using specialised tools, but an understanding of the standard tools used is needed for designing the systems in use now and for the immediate future.
When examining a void like this it may be necessary to look both to the latest research in the fields in question, and to older work which may fall somewhat outside the thinking that led to the gap existing in the first place. That path leads down a road through the works of cognitive scientists such as Tarja Susi (2005, 2006), sociologists such as Barbara Rogoff (2003), psychologists like James J. Gibson (1986) and Eleanor Gibson (Adolph et al., 1993), all the way back to philosophers and phenomenologists such as Maurice Merleau-Ponty (1945/2005). Retracing some of the thoughts that got us to this point may point to hitherto overlooked approaches to the problem.

Much research can be found on human learning and skill acquisition. Examples of this include Rogoff’s (2003) book on the cultural nature of human development which approaches learning and behaviour from a social psychology standpoint. Tool use and the extension of the human body through tool use has also been extensively covered from various research perspectives, with the already mentioned embodied cognition standpoint (Susi, 2006) being of particular interest for the purposes of this thesis.

Computers and virtual interfaces have also received much interest in recent years, with focus being put on how we perceive virtual environments in relation to our own body (Geuss, Stefanucci, Creem-Regehr & Thompson, 2010), and how we perceive the possible actions in program interfaces (Gaver, 1991). A popular concept used within research on virtual environments is affordance, which is a term used to explain how the actor (in this case, computer user) perceives and acts upon action possibilities (the actual possibilities for action allowed by a virtual object).

The articles mentioned in the various fields, tool use research, motor-skill development, child psychology, HCI, each seem to use one of three types of explanations from within embodied/situated cognitive science for how actions are made possible. These explanations are: high-level cognition and tool use, affordances (direct perception or mental model based), and action-perception coupling (also referred to as action-perception loops). These particular views emerged as being relevant through the search patterns performed, that is, searches for articles on how an understanding of the use of a two dimensional tool in a three dimensional environment led to articles that focused on one or more of those approaches for explanation. All three concepts are embodied and situated in nature as they veer away from the older paradigm of the mind as a computer and separation of mind from its surrounding environment, opting instead to explain our understanding and interaction with the world through viewing the body and environment as a part of our cognitive processes. These approaches can be seen as part of a single approach, yet do have differences in focus, which is the reason for them being used here. The high level analysis will focus on socio-cultural aspects, internalised knowledge and knowledge learned from others (Susi & Rambusch, 2008; Lindblom & Susi, in press), through various mechanisms such as mediated learning and learning by example (Rogoff, 2003). That high level analysis then also involves using internalised knowledge learned for other domains, and applying it to the domain being examined. The affordance analysis is an approach based on information flow from the environment of actions possible to the user, and has different interpretations with very different models of
picking that information up and acting on it. The analysis using the perception-action loop concept will approach the task through a more mechanistic approach which examines how an action can be completed through a relatively simple loop of input/output, perception/action, with only a goal state in mind but without the internal processing of expected outcomes of each part of an action, thus focusing on the connection made by the action itself (Clark, 1997).

The goal in this thesis is thus to explore in more detail how existing research may explain what is required to perceive and act upon action possibilities in a virtual environment. This is done with the hope that a better understanding can be achieved from which it would be possible to perform further research for supporting research and development within human–computer interaction (HCI), giving a chance for developing models for action-perception between virtual and real environments, as well as guidelines for user interface (UI) design and development based on grounded theories. The final results might be that this gap can be easily bridged with current theories of tool use or other well known theories, or it might be discovered that the jump from real into virtual is too large for current theories to cope with when there is a mismatch between the virtual interaction and the tool being used for interaction. Any answer, even a partial answer or something from which an answer may eventually found, would be useful for the fields of cognitive science, HCI, and UI development. That said, it is quite possible that any results may contribute more to one of the mentioned fields than the others.

2.1. Methodology of the analysis

The analysis chapter will analyse a hypothetical use scenario based on the context presented in the previous chapter and illustrated in Figures 1 through 3. Unless specifically noted otherwise this will be done by the author using personas, which is a method for using a made up character to role-play through a scenario such as this. The analysis will be done multiple times based on the main themes identified in the prior research which has been presented. The analysis will first be done from a standpoint with a focus on high level cognitive functions (abstract, internal, socio-cultural), then from an affordance based standpoint (context, the environment), and finally from a perception-action loop perspective (low level, senso-motoric). All these analyses come together to form a complete situated/embodied approach, but the three approaches give somewhat different viewpoints from which to find, and hopefully explain, where the cognitive and action gap causing the picking and drag-threshold problems exists, why that gap exist, and how it can be bridged. Affordances are by their definition very contextual, relying on the on an object’s placement in its environment. Affordances thus rely on the environment and the perception thereof, and if Gibson’s (1985) original definition is used then feedback from the world is only partly explained through changes in viewpoint on the object which gives off the affordance. Pure senso-motoric explanations are not well supported in Gibson’s (1985) original definition, as he focuses on the visual system and only explores interaction through the movement of the observer, not through the manipulation of the object being examined. Action-perception loops are,
conversely, senso-motoric in their nature, dealing with the feedback caused by an action.

The three analyses are used to explore different aspects of the scenario so that the nature of the gap being examined may be determined. This is somewhat akin to a person using multiple ways to learn about an object such as looking at the object, lifting/touching the object, and smelling the object to determine its nature. With only partial information, for instance only visual information, it may be impossible to determine whether an object is made made from granite or papier-mâché. Not fully understanding the nature of the object/problem can lead to wrong deductions as to what can be done with the object/problem. An object made from papier-mâché may not be good for building with, and an analysis that focuses only on the contextual, or affordances, may lack an understanding of how the action is learned through doing or any abstract knowledge possibly used to make sense of inconsistencies in behaviour. Anything built from those false premises will likely crumble when put to use in the real world. Conversely, performing an analysis using all three approaches rolled into one might give the same information as the three-pronged approach, but pinpointing the nature of the problem would be harder due to all the information arriving simultaneously, which would complicate crafting a solution, a bridge. Finding out exactly which parts of the scenario are supported by each part of the analysis method is crucial, and therefore a three part analysis becomes a simpler solution which should give data that is easier to work with further. In a best case scenario that would result in the high level analysis giving information about how a previously learned understanding helps us complete the scenario, the affordance analyses would explain what information the environment itself gives off during each part of the scenario, and the perception action coupling would give information about how the scenario can be completed through only modifications of the movements perceived without the need for any abstract concepts. Examining the movement of the mouse and cursor and attempting to analyse separately how the jump from layer 2 to layers 3 and 4 can be explained by prior understanding, the perception of affordances, and a loop of perception-action is thus used to see whether any one of these analyses sufficiently explains the jump between the layers on its own. If one approach (such as affordance) is sufficient to properly explain the layer jump then the problem has been solved, but if none of the approaches explains the layer jump sufficiently then it is possible to try combining the results from all three analyses. If this does not give a full explanation of how a person can bridge the gap between layers then either the selected methods are not useful for this kind of analysis, which would cast some doubt on their overall usefulness for the analysis of actions, or that the analysis approaches were not used correctly, which in turn would raise the question whether the analysis approaches are sufficiently clear for being used this way for analysing actions or whether the blame would lie with the author.

The results from these analyses would then give information about the components of the movement and perception involved in the action, information about how possibilities for action gets perceived, and how all that gets added to more complex models of understanding. This can then be combined at the end to see how the action can be learned through doing,
through sampling information and how the action can then be repeated. Any specific areas where a theory breaks down and cannot explain some part of the action can then be marked as a possible problem area, and if multiple analyses approaches fail to provide usable explanations for the same part of the action then a problem of some sort has been positively identified.

The final analysis thus aims to use the results of three analyses using tools which are related but different in focus, and compare the results of those analyses to find an explanation for how actions and perceptions are transmitted from layer 4 to the other layers. Any explanation found will then have a body of work behind it from the analysis theory which supported that explanation, which can then be used for continuing research and design on three dimensional computer environments using standard mouse and cursor driven interfaces.

2.2. The scenario defined

The scenario involves a three dimensional computer drawing program in which it is possible to create basic shapes and combine to form more complicated objects. A hypothetical user which represents an experienced user of this particular system, i.e., knows what an action will accomplish and is not learning the tasks, will be used for performing the actions. Additionally, some parts of the actions will be examined through the eyes of of a novice who does not know the actions and translations between layers. Which hypothetical user is being used at any given time will be specified. The reason for using two different hypothetical users is that the more experienced user should have most processes for performing the actions already internalised, which allows for analysis of action based on understanding, while the novice user persona will allow for an analysis which shows how exploration of the action and learning occurs. Using personas with different basic approaches is done in the hope that it gives the three analysis methods somewhat different information with which to work, as the novice persona should give off more information which can be explored through affordances and action-perception loop theories, while the higher internalisation of the expert user should result in more information that can be viewed as high level and abstract. This is due to the analyses using this persona approach and having access to the internal thought processes of these virtual test subjects.

This approach is in line with scenario based design and evaluation, and borrows much from the scenario and persona approaches to design, in which a hypothetical user is specified and a possible use scenario explored using that persona to find flaws in a system (Benyon et al., 2005 and Cooper et al., 2007). This approach is especially useful when the product does not yet exist, or when experiments are not feasible (Cooper et al., 2007). There are various approaches to the actual design of the persona, but all require understanding the needs of the product and/or scenario. In this case the persona(s) can be mostly based on the researcher, and the definitions that are required are mostly to do with experience level (how much the user has to learn/explore).

The user is working in a three dimensional object building program, see Figure 3 for reference. He has created two cubes, and needs to line the two up
horizontally to combine them into a new shape. The user moves the mouse to move the cursor, which moves left/right and up/down on the screen. He moves the cursor so that it is directly in front of or “on” the rear cube (numbered 2 in Figures 12 and 13), links the cursor to the cube by pressing the mouse button, and makes sure the mode of object movement is set to the z axis (the “away/towards” axis), although the cursor will continue moving in the x and y axes, and drags the mouse towards himself which causes the cursor to move along the y axis but causes the cube to move along the z axis. Note that the selection of axes did not come into play during the object selection action, only during object manipulation. Furthermore, the mode switching can be done through different mechanisms such as the previously introduces three separate 2d views, mode switching through button presses, mode switching through tool selection in the 2d interface, or by using the drag-threshold model used by Google Sketchup and AutoCAD. This shows how the scenario is designed for separate analysis of two known problems in the task, the picking problem (object selection) and the three dimensional movement problem (such as the drag-threshold problem).

The task is kept extremely simple on purpose, but a question then arises to whether the scenario can be seen as being representative of real world tasks in the domain? Representing a typical task might be seen as being tricky, but it becomes important to understand which action (or even sub-action) is likely to yield information on the problem at hand. As the gap being examined involves the problems of selecting an object and moving an object it is enough to create a task that involves these two known problems and the core distractions which make those problems more difficult to solve. The picking problem is represented by selecting a three dimensional object in three dimensional space, with the distraction of a second object placed closer to the viewer and possibly partially occluding the object to be selected (based on viewpoint). The drag threshold problem is represented by moving the object in a different plane (depth) to what the cursor is moving (up-down, left-right) and possibly not even in a completely vertical-to-depth translation, that is, there may be a horizontal factor in either movement that is not necessarily translated between the two dimensional and three dimensional domains.

Figure 12: Two cubes in a program with cube 2 placed further “back”. Figure 13: Cube 2 has now been moved “forward” to sit adjacent to cube 1.
The simplicity of the scenario is intentional for examining the identified gap in theory; if the scenario were more complicated it is unclear whether any more information would be gained but the cost would be a much bigger analysis which would increase the chance of overlooking the important jump in action and perception from two dimensions to three dimensions. Examples of more complex scenarios are the creation of more complex geometry such as extrusions along curves, the combination of objects, and rotation of objects. All those scenarios are representative of tasks performed in CAD applications, but share difficulties with the simple scenario used here. These difficulties mostly have to do with precision of movement in three dimensions, understanding along which axis an object will move or rotate, and understanding how a two dimensional representation will convert into a three dimensional object through extrusion. The latter is yet another problem of understanding how a motion along different axes will show up on the screen. All these scenarios are more complex to visualise and analyse but appear unlikely to give any more useful information and therefore it was decided to use the simplest scenario which should yield information about the problem at hand, the gap between layers 2 and 4. This simple scenario was designed to give the shortest analyses possible which actually contains the problem area which is being examined.

2.2.1. Personas

Personas are a tool used often used by usability experts for evaluating systems when access to the system is not available or not likely to be effective for some reason (Cooper et al., 2007). A persona is a description of a user which is considered to be likely to give information about how well the system works through observation of his actions. The expert then mentally goes through a scenario and tries to behave in a manner consistent with how the character created in the persona would likely behave, and make decisions based on the persona. The level of detail can vary, but incorporating any characteristics considered likely to affect the use of the product being analysed is of primary importance. Using a persona is thus a role-playing of sorts.

Two personas are used in this analysis, one with CAD experience (expert persona) and one with no CAD experience (novice persona). The expert persona is roughly defined as having prior experience of the domain (three dimensional building programs such as CAD programs), and high general computer experience. Characteristics of the personas that are not specified can be considered to be based on the expert performing the analysis; a male in his mid thirties with no impairment for interacting with a computer.

Expert persona has high general computer skills and high domain skills:

- Using a mouse is not a problem
- The interface in layer 2 is not a problem
- Performing actions in layer 4 is something the user has experience with and is not a problem
The less experienced persona, which will be referred to as Novice, has high general computer skills, but little domain skills:

- Using a mouse is not a problem
- The interface in layer 2 is not a problem
- Performing actions in layer 4 is something the user has NO experience with and may require exploration and learning

These definitions are done so that the focus can remain on the gap being explored, the jump from layer 2 to layer 4. Lower computer experience would shift the focus towards the interaction of the user with layers 1 and 2, as the user would have to figure all those interactions out to begin with. This results in even the novice persona having the knowledge required to use the link between the mouse and the cursor, as that link is not the focus of this thesis, as well as this level of knowledge being more the norm in today’s society. The reason for using a persona rather than using test subjects is that what is being explored is an internal thought process, and should thus be easier to understand and document if performed by a usability expert (Cooper et al., 2007).
3. Analysis

This chapter will go through the three analyses in order and will end by comparing the three analyses. The analyses attempt to find an answer to the question how two dimensional interaction methods function in a simulated three dimensional environment, and does so by using three different approaches to analysis.

The scenario was analysed from three perspectives, high level, affordance, and perception-action loop. The task itself was broken into action and perception components as shown in Figure 14. As there is no physical feedback through the system then all lines pointing right (and coloured black) are actions, and all lines pointing left (coloured blue) are perceptions. Vertical arrows connecting to perception arrows from layer two elements indicate a element for perception from layer two. Vertical arrows (green) in layer 1 show where a perception must be converted to an action. The analyses were performed by the author playing the role of the user, through the use of personas, and mentally performing all steps of the scenario, using figure 14 for reference and assistance. The different perspectives were kept in mind during each performance of the scenario and each motion viewed with the question “what is going on from the view of the currently used perspective” in mind. The scenario was also performed in Google Sketchup to verify that the scenario itself was indeed realistic, even if it uses only simple objects and motions. A real user was brought in to verify whether the persona based analysis was accurate when a surprising behaviour was found in the perception-action loop analysis when using the novice persona. This real user was asked to perform the scenario in Google Sketchup as she did not have the required experience to do the scenario as a scenario based thought experiment. That real user’s experience level closely matched the novice persona’s experience level. Each analysis in turn went through the scenario using figure 14 as a guide to examine what processes were being performed at each step in relation to the analysis perspective being used.

3.1. Analysis from a high level cognition perspective

The high level standpoint is excellent for describing the actions in an abstract way, and eschews the low-level mechanical description of perception-action coupling, as well as the the contextual nature of affordances, focusing instead on a general understanding of the actions in easily readable terms.

The default persona had a prior understanding of computer systems in general, and this domain/program in particular. The prior understanding of computers in general means that he had an expectation as to how moving the mouse would affect the cursor, as well as expecting items which appeared to line up to offer possibilities of interaction. His prior experience with this particular domain and program also meant that he expected the cursor to permit linking the mouse to any objects within layer 4 which the mouse could line up with.

The novice persona on the other hand only knew for certain how the movement of the mouse would affect the cursor. He had no knowledge of how
the cursor would interact with the cube, as his usual approach of reaching his hand to the object and creating a link to it by grasping with his hand is impossible and must be substituted with a tool which cannot perform the three dimensional reaching and grasping action in a similar manner. His experience with real world objects suggested that using his hand or any available tool on an object would be likely to cause an effect on that object. As he was already aware of how the mouse affects the cursor and already viewed the cursor as a tool within the computer environment he was likely to try using the cursor to manipulate the cube. He was likely to try sliding the cursor over the cube (over the line of sight made by the forced perspective), as a reaching/grasping movement would have made the hand align visually with the object, and when the cursor itself did not affect the cube he would have tried pressing the mouse button. He tried the left mouse button first, as that is the button which causes the cursor to “press” onscreen buttons, and when he did that any movement of the mouse will have revealed that some kind of link has been created between the cursor and the cube. The novice had computer experience and so was already accustomed to selected objects being movable in some way, e.g. icons on the desktop or text in a word processor. The object had now become movable, with the motions sometimes matching the user’s motions, and in other cases moving in a spatially different, but visually quite close, direction. This is where the novice’s expectations through embodiment broke down. He could no longer trust the motions of his hand to be simply translated to the virtual environment “within” the computer screen (either as

![Diagram](image-url)

**Figure 14**: The scenario has been split into components, by layers. Arrows pointing right (black) indicate actions, arrows pointing left or vertical in layer 2 (blue) indicate perception, and vertical arrows in layer 1 (green) indicate a perception being converted into an action.
a simple vertical motion, as the cursor, or a to/from movement as the mouse) even though the cursor still displayed this simple link. The changing directions of the movements of the simulated object’s movements broke down the novice’s ability to rely on an embodiment or body schema based mapping between the layers. Down was no longer Down, and towards the user was no longer towards the user, as they would shift with minute movements of the mouse to the left or right. Left or right movements had suddenly become the selector for whether the total movement of the simulated object was down in the virtual environment, or towards the user in the virtual environment. The novice then tried moving the mouse cursor to the place where he wanted the simulated object to be placed, and when that sent the simulated object down in the virtual environment instead of towards him he tried moving the cursor sideways as that had previously caused the simulated object to jump between places. This caused the object to be positioned close to the user’s viewpoint, and close to his intended goal.

This approach to the analysis gives an overview of the actions, and discusses the personas’ understanding of the system, but does this in a very abstract way. This results in any explanations given being quite vague and not very useful for the development of a model of the action which would give tools for system development. The high-level explanation of the actions shows an interesting viewpoint, but requires much more explanation of any underlying mechanisms of cognition to be truly useful. This high level of abstraction is one of the strengths of this kind of analysis as it gives an easily understandable form of results, but the resulting high level explanation may itself require analysis for understanding the rationale behind each part of the results. This is because the high abstraction level of the explanation lacks a model for explaining the processes used for achieving the desired action.

3.2. Analysis from an affordance viewpoint

To begin with, the user saw the pointing device which afforded sliding around on the surface on which it lay. If the mouse was examined in more detail then items such as slippery pads on the underside would have become apparent. This examination is supported by Gibson’s (1986) view on sampling of information and physical movement being a necessary part of his concept of affordances through direct perception. The mouse buttons afford pressing which aligns the hand naturally on top of the device, and the cord which connects the mouse to the rest of the computer would also have given information as to the intended range of motion as well as the intended orientation of the device. A mousemat (if used) gives further information to what plane and size of motion are required. This enhances the sliding affordance of the mouse by giving an additional channel of information to that provided by the mouse itself, as well as suggesting that the mouse should be used on this horizontal surface and not directly on the screen.

Coming to the cursor shown on the computer screen a question arose: does the cursor actually have or give off any affordance information until the user has moved the mouse itself and established a connection between the two? This depends on which definition of affordance is being used. Gibson’s (1986) direct affordance can be seen to have a problem with this, as an arrow shape
on a computer screen cannot easily be seen to be connected to the physical mouse. Note that this is the case for other objects where no direct coupling can be seen between tool and object, such as remote controls and televisions. After a causal link has been established it becomes easier to see an invisible link from the mouse to the cursor, and this movement could be explained as a detailed examination of the object as per Gibson’s (1986) theory, with the invisible link being made visible through this movement of the object. Gibson (1986) mostly describes the movement of the person perceiving, but does not rule out manipulation of the object being examined if that is what is needed to sample the required affordance information. This is then the level of complexity that exists between layers 1 and 2 (see Figure 2, pg. 9). Note that Gibson (1986) does not exclude learning from affordance, and in fact Gibson mentions learning, but he does not give any explanation as to how learning might be supported within his theory of direct perception of affordances.

If Norman’s (1986) mental model based affordance is being used then the problem is simpler to solve and is based on previous experience or other forms of learning to see the mouse and the cursor as being conceptually linked. Watching someone perform an action, or moving the mouse and seeing the cursor move in response, adds an understanding to the user’s mental model that the mouse affects the cursor. The cursor gains the affordance of being movable and connected to the mouse through learning from others with more experience (Rogoff, 2003), or through experimentation. Now that a connection has been discovered between the mouse and the cursor the user can move the cursor. The cursor has thus become a tool, an extension of the mouse. The onscreen controls in layer 2 (the two dimensional interface layer) tries to show similar affordances to real buttons and has selectable areas. These affordances suggest pressability, but pressing the glass of the screen with a finger does nothing; the buttons are as if protected by a sheet of glass. As the user has found a tool which lets him perform movement behind the glass we now see a phenomenon called nested affordances. Nested affordances are those affordances that rely on other affordances to be possible; utilising the affordance of pulling on a door handle does not become possible until the door handle has been rotated downward (Gaver, 1991). Until that configuration of handle had been achieved the “pullable” affordance was a false one as it suggested an action which was impossible to perform without first performing another action. Pressing an onscreen button only becomes a real affordance when using the correct tool, the mouse. Until that point there was only the misperception of a false affordance (Gibson, 1986) for pressing on the screen’s surface and the correct affordance had not been discovered. This functions well with Gibson’s (1986) theory of affordance as his definition of affordances centres on the actions made possible to the animal in question (user in this case). By using the mouse the user has effectively changed what actions are possible for him, and gained access to other affordances which were previously unavailable to him. The user can now interact with the two dimensional controls (layer 2) with impunity, and the buttons on the mouse provide a similar affordance (pressing) to those on the screen, so it is no big stretch to see the tool (the mouse and the cursor) as extending the possibility for pressing from layer 1 into layer 2.
The user wanted to line up the two cubes in layer 4 as shown in Figures 7 and 8, and they (the cubes) should have been giving off similar affordances to cubes in the real world (layer 1) regarding rotatability, movability, stackability and so forth. In fact, the cubes can be seen to be visually analogous to two cubes behind a glass pane. Information is thus given about what actions may be possible with the cubes, but the cursor does not give any visual information suggesting that it can reach “into” the screen. The cursor only moves left/right and up/down, yet even cube number 1 (see Figures 7 and 8) is clearly represented as not being in the same plane as the objects shown in layer 2 (see Figure 2). To explain this the direct perception approach to affordance needs to go back to the explanation that was used for bridging the gap between layers 1 and 2, namely that the user must discover the possibility for linking the cursor to the cube through experimentation, instruction or observation. That would suggest that in the absence of instruction or observation of another user performing the action the novice would either click around randomly, or try clicking “on” the cube and then moving the mouse around randomly until something happened. Using the novice persona I see other, more directed actions appear. As that was somewhat unexpected an actual novice user who had never before used a CAD program before but did have decent general computer skills was asked to try to perform similar actions in Google Sketchup. The observation which was made by using the persona was supported by observing the actual novice user; tasks were not random but rather exploratory in nature. This exploration seemed based on expectations of behaviour and logic based on feedback from the system. Problems may occur when attempting to move an object on each of the three axes of possible movement, but each motion of the mouse is directed with purpose and intent, even when the cube does not obey that intent perfectly on the first try. When the cube did not move as expected, the real novice user was observed to move the cube back to its point of origin and try again with subtle differences in the motion, which was identical to the behaviour predicted when the persona was used. This can be seen to support the accuracy of the persona, showing the persona to give a good representation of an actual user.

Affordance thus has a problem explaining this action if only the direct perception approach without internal representation (Gibson, 1986) is used as an understanding of the basic mechanics of the computer domain seem to be used as a building block for use in the specialised CAD domain. This suggests that more is needed than only direct perception. If, however, Norman’s (1986) mental model based view on affordance is used then an element of high level cognition has been added. Prior experience and understanding of the domain, and possibly across domains, gets added to the concept of affordances which then integrates high level abstract cognitive ways of understanding the world with the more context based concept of affordance. This high level understanding of the relationship of the mouse to the computer, and the simulated objects within, allows the user to base his/her actions on prior understanding from other tasks done in the computer domain. This would apply to all forms of learning where learning can be seen to use any previous understanding which is to be applied to a new task or domain. Tasks such as dragging and dropping files, moving objects within two dimensional interfaces and selecting text all give certain models for
expectations of movement. This explanation, however, lacks detail and lacks a model or detailed explanation of how the necessary new knowledge for moving objects in three dimensions then gets added to the existing mental models for motion in two dimensions. It does not explain the action, but only states that it can support the action. More is required to make this approach useful for task analysis.

An affordance-based view based on the works of philosopher Maurice Merleau-Ponty has also been suggested (Dohn, 2009). This approach shares more in common with an embodied and situated viewpoint as it revolves around the social and physical interactions of the user with the environment and how those interactions relate to the users body schema. Body schema refers to how actions’ and concepts’ relation to our bodies are the basis for our understanding of actions, our understanding of concepts, and how artefacts can be incorporated into our sense of ourselves (Merleau-Ponty, 1945/2005). This suggests that up-down, left-right, and toward-from, above-below all need to be interpreted in terms of their relation to the body. Dohn (2009) suggests that a body schema view is essential for making sense of three dimensional interaction on a two dimensional screen, but does not go into specifics which show whether this view can be used to explain the gap between layers 2 and 4 (as shown in Figure 2, pg. 9). The problem with using this to explain the transition from layer 2 to layer 4 is that although the screen becomes an extension of the user’s eyes which allows him to see into an environment which would otherwise be invisible or non-existent, and the mouse links to the cursor and thus extends the user’s possibilities for action, there remains a lack of an obvious extension in both directions from the two dimensional to three dimensional due to how the functionality of the mouse/cursor combination rapidly changes from two dimensional interaction to three dimensional interaction, and within the three dimensional interaction changes between x/y axis translation and x/z axis translation. These changes in functionality mean that even suggesting a simple extension through the layers, like I have suggested between layers 1 and 2, has problems as this invisible link is quite complicated and requires abstract thought and understanding which does not integrate easily with a purely body schema based interpretation of affordance due to the multiple and almost randomly changing functions of the tool itself within the complex virtual three dimensional environment. Body schema, as was explained earlier, relies on an understanding on the world in relation to the body. Performing actions with a mouse limits the body’s movement to two dimensions, and thus any body schema for interacting with the computer can be seen to use the limitation of the current movement for reference. This results in two dimensional movements on the computer screen being quite simple to explain, as they match the currently active body schema, and two dimensional manipulation of simulated three dimensional objects can be explained through the link between the cursor and the simulated three dimensional object while the object moves along one set of x,y coordinates as there is a direct mapping between the movements. The abstract thought, and problems with a purely body schema based view, come into play when the two dimensional movement of the hand and mouse suddenly start applying to another set of x,y coordinates without any visible indicator of the system having changed modes and no change in the movement of the hand and
mouse. This requires a complex understanding of how the simulated object in the virtual x,y,z environment has rules that govern its motion which do not conform to how motion works in the real world, on which the body schema view is based. Moving an object towards oneself in the real world will always result in the object moving towards oneself, while in a virtual environment using a mouse and a drag threshold system for interaction an object may move towards the user or downwards in the x,y,z environment based on largely arbitrary rules. These rules are generally selected to align with the X, Y and Z axes so that a motion of the cursor which is close to the visible alignment of one axis or another will make the simulated object move along that axis.

The problem arises when we consider that the two dimensional action of the hand/mouse is inconsistent with the varying nature of the three dimensional manipulation of the simulated three dimensional object. Looking back to the Flatland (Abbot, 1884/1992) example, the two dimensional being cannot even see or imagine the third dimension as the space for that does not exist in his euclidian world. Once the user has limited his hand movements to two dimensions it becomes hard to use a body schema to explain the three dimensional movement of the virtual object. There is no problem in understanding how a body schema can be used to relate the simulated objects to one another, or to explain the user’s understanding of depth or location in reference to himself. The only problem area is the sudden break in dimensionality between the motions of the hand and the tool on the one hand, and the simulated object on the other hand. The body schema itself is three dimensional as it refers to points being left/right, above/below, or in front of/behind a person. The problem only comes into existence when there is a discrepancy between dimensionality of the object being referred to or manipulated and the tool being used to manipulate it.

The correct action to continue at this point would be to place the cursor so that the forced perspective makes it appear to be in line with cube number two (see Figure 7), press the mouse button to link the tool (the cursor) to the object (the cube), and then drag the mouse towards the user causing the cursor to move downwards and the cube to move towards the user.

There is no visual information to suggest these actions, so something is clearly missing from Gibson’s (1986) original direct perception version of affordance. A Merleau-Pontian (Dohn, 2009) view incorporating body schema could explain how the action is performed when there is a dimensional match between the interaction tool and the virtual world, but has problems with the sudden break in dimensionality between layer 2 on the one hand and layers 3 and 4 on the other. This will be examined further in chapter 3.4. Norman’s (1986) mental model version of affordance can explain the action through viewing the gradual buildup of an understanding of the system starting with the first step (discovering that the mouse moves the cursor). This explanation requires trial and error as a part of the building of the mental model, which does not fully correspond to personal observations of inexperienced computer users grappling with new use cases such as the one used for analysis. The trial and error part will be examined in more detail in the perception-action coupling analysis.
3.3. Analysis using perception-action coupling

This approach to analysing the scenario is in some respects simpler than the preceding analyses as it is purely senso-motoric in nature, and eschews abstract processing of internal representations or high level concept of prior knowledge to successfully complete the task. It is important to note that perception-action coupling can involve some forms of internal representation, and indeed it would be possible to argue that Rizzolatti and Craighero’s (2004) work on mirror neurons shows exactly that. At a more basic level, any object observed by the sensory apparatus will trigger some pattern of brain activation. That says only that the eyes, ears, and other sense organs are connected to the brain, not that there is abstract processing of the sense stimuli going on. Action-perception coupling is thus a way of working with and explaining perception and actions without abstract thought and internal manipulation of objects being required for basic functioning (Gaver, 1991; Clark, 1997).

Discovering the link between the mouse and the cursor (without any prior experience) requires nothing more than experimentation. The user presses the onscreen button directly with his finger with no effect (an action which causes no change in the environment). Any action that moves the mouse, whether on purpose or by accident causes a perception action loop to become active, with any movement of the mouse causing a change in location of the cursor, which again allows the user to modify his movement. This same behaviour could bridge the gap between layers 2 and 4, as the novice persona sees the object he needs to move, tries moving the cursor over the cube and sees whether the cursor extends towards the cube or causes it to move, tries pressing the mouse buttons (seeing as how the mouse has already been seen to connect to the cursor). When these actions do not cause any movements of the cube he ends up combining the two actions he has discovered that the mouse allows him to perform. Thus the novice ends up with the cursor visually lined up with the cube, and tries pressing a button and moving the mouse at the same time. This causes a movement which the user perceives, and can then try movements in different directions until a movement is found that brings the cube closer to the desired end state. Small changes in movement may change the axis of movement, but the user perceives instantly that the movement is no longer bringing the cube closer to the desired goal and modifies the movement accordingly. When those movements have been discovered and the cube placed in the desired end state the novice must disconnect the cursor from the cube, which is done in this example by releasing the mouse button.

The default persona however does not need to do any of this exploring of possibilities. He grabs the mouse, moves it and sees the cursor move to the cube, presses the mouse button and drags the cube to the desired position, quickly correcting any mistakes in movement if the cube does not move correctly towards the end goal. Interestingly, the novice user who was recruited to test the scenario did not show this exploratory behaviour either. This suggests that perception-action coupling must be used in conjunction with another analysis which works at a higher level of cognition and incorporates some model for applying skills learned for use in a different domain in this one.
Perception-action couplings work well to explain the discovery of action possibilities and the learning of actions (Lockman, 2000), but does not explain more complex elements of action such as the prediction of outcomes after an action has been learned as the expectation of outcomes requires more abstract and higher level cognitive work using concepts such as representations (Susi, 2006). Perception-action couplings also allow for an explanation of complex actions without requiring any internal representations of the tool/object such as mental models (Susi, 2006). This is effectively why perception-action couplings are useful for examining how action possibilities may be discovered, and works examining tool use by adults, infants as well as non-humans such as chimpanzees or crows (Susi, 2006). An obvious downside to this approach is that it is imperfect both for examining those who already have experience with the action and, more importantly, why a novice’s actions use other methods than pure exploratory methods. Using the perception-action loop concept in conjunction with some form of internal representation concept should solve this.

3.4. Comparison of analyses

The preceding analyses show that the different approaches to analysing this scenario each have their strengths and weaknesses. One thing that is lacking is a reliable way for discovering how the simulated three dimensional object can be moved without resorting to randomly testing motions. Randomly pressing buttons is not what is observed in new users of such systems. This was discovered during the persona based scenario analysis and verified by informally observing one novice who performing a version of the scenario which was the focus of the analyses. There is a clear prediction by even the novice that pressing the mouse button while the cursor hovers in front of an object and then moving the mouse around should result in some kind of motion, and the three dimensional movements were directed in nature with the novice showing clear expectations as to what should happen. This must be based on some kind of high-level conceptual knowledge or understanding, which should be easy to add in from the high-level analysis, but the information resulting from that analysis does not easily avail itself to providing a clear explanation as to exactly what this high-level conceptual knowledge consists of. In fact, the high level analysis seems to be missing the crucial step of actually explaining itself. That whole analysis goes from theory to results, seemingly without any clear definition of how it actually works and gives those results.

To successfully incorporate a high-level element into the analysis requires a more concrete model, especially to deal with scenarios like this one which contains a break in the dimensional continuity of the action/perception.

Lockman (2000) suggests that tool use must be seen from the point of view of relations between elements of that object, relations between objects, and objects relation to the environment. I would contend that combining that view with a Merleau-Pontian body schema view results in something that closely resembles an embodied and situated viewpoint, which is not at all surprising seeing as how Merleau-Ponty is referred to by Susi (2005 and 2006), Clark (1997), Rizzolatti and Craighero (2004) as well as others which are referred to in this thesis, all of whom support an embodied view of cognition. This still
points back to the break in dimensional continuity, as the relation of objects adhering to differing dimensional rules seems complicated, and body-schema approaches break down when first limiting motion to two dimensions and then trying to elevate that motion to three dimensions again. Body-schema relates objects, movements, and positions to the body and its capabilities which means that tool use wherein a tool affects links to an object or manipulates it in some way (such as when grabbing something with a pair of pliers) is easily seen as a simple linear extension of the arm, giving a set of advantages (strong grip through mechanical leverage, less heat sensitivity) and imposing some limitations (no surface pressure sensors, less dexterity). The jump from layer 1 to layer 2 is likewise simple. A simple link which is close to being a one-to-one link in movement is made, which gives the advantage of being able to manipulate objects that can be seen but not touched inside the monitor, but imposes the limitations of lack of feedback and of two dimensionality. These limitations are imposed through both the limitation in movement of the mouse itself, and through the visual feedback from the cursor which moves only in two dimensions. This type of limitation is nothing new to the body-schema. Any object which can only be slid along the ground, for instance. Cars, even toy cars, share this limitation. Even children learn to follow these limitations; a toy car only slides along the ground unless the child is playing with having the car change into something which does not have these limitations such as an airplane. What the body schema is not set up to do, I contend, is to limit itself to one paradigm of action such as a two dimensional functioning, yet have the results suddenly be three dimensional. This results in confusion when the dimensionality of movement is not stable, as is shown in figures 10 and 11 on pg. 13. The gap between layer 2 on the one hand and layers 3 and 4 on the other is thus caused by the lack of a compatible frame of reference between the two dimensional layer and the three dimensional layers.
4. Conclusions

This thesis aimed to examine how a two dimensional controller can support action in a simulated three dimensional world both from a cognitive and a practical view through a study of existing literature and an analysis based on making a hypothetical scenario which should be representative of a real world task which would then be completed by the author of this thesis using defined personas to simulate users. This has only been partially successful, with more results for practical design of three dimensional interfaces than for the cognitive underpinnings of understanding how a person manages to work in current three dimensional programs using two dimensional tools. That said, some progress was made towards a deeper understanding of the problem through the use of the body-schema concept, as discussed in the previous chapter.

The conclusions are that the analyses methods used proved ill suited to answering the question of how the action traverses the gap between layer 2 and layers 3 and 4, but the analyses did pinpoint the problem itself quite well, showing that it did indeed exist at the point where two dimensional movement translates to three dimensional movement. Understanding what the problem is, where the problem occurs, and how it can be circumvented, has practical applications as that understanding can be used to form design guidelines which assist designers of three dimensional applications in avoiding the use of interaction and interface elements which have the identified problems and would therefore be unnecessarily difficult and inefficient to use. A set of simple guidelines for the design of three dimensional interfaces for programs such as CAD programs will be introduced later in this chapter. The analyses methods proved ill suited to the task at hand which seems obvious in retrospect as they are ill defined and require much interpretation. Additionally, the author may have been somewhat careless in his original

Figure 15: A four panel interface for CAD programs. Each two dimensional panel allows for interaction with two axes, and the fourth window displays the resulting movement of the complete object in simulated three dimensions.
interpretation, particularly of the high level analysis where a confusion of terms may have caused that analysis to be less useful than it otherwise might have been. All that said, the literature study and the selected analysis methods did turn up something interesting by pointing the author towards the body-schema concept. Body-schema provided an angle for interpreting why there was a breakdown in interaction between the two dimensional layer and the three dimensional layers.

The problem with the main analysis methods was that the gap between layers 2 and 4 was only bridged by a rather vague type of the user “having knowledge” of what was supposed to happen. This is clearly not what the concept of affordance is about according to Gibson (1986), so the affordance concept was not successful in explaining the scenario on its own. Neither does it match regular analyses of action through tool use, as regular tool use has all perception and actions occurring in an equal number of dimensions. The more mechanical approach of perception-action loops can easily explain the actions through the use of iterative actions which bring the user closer to the goal but requires major additions to explain the directed and intentional parts of the movements, and affordances suffer from being a group of unconnected and contradictory approaches sharing an ill defined concept. The best affordance based explanation came from Norman’s (1986) affordances, but this theory suffers from a lack of clarity which means that no concrete explanation emerges which can be further built upon. This is where the action possibility concept gets useful as it can be separated from the mechanics proposed in the various affordance theories. Action possibilities could thus be worked into the other theories, or used as a module of sorts to describe only the the actions possible with a certain object, which can then be examined using other theories more suited to task analysis.

Examining the gap through the use of the body-schema concept turned up the most surprising answer of all; a possible answer to what causes the problem between the two dimensional and three dimensional layers. The lack of a compatible frame of reference when attempting to perform a three dimensional movement after first having limited the tool (and hand) movement to two dimensions seems plausible, and would explain the problem’s cognitive roots. The problem thus seems to be the mismatch in visual information given, which appears to be three dimensional, and the interaction method provided, which is two dimensional. The feedback of the system occurs in the simulated three dimensional environment and is always limited to two dimensions at a time and uses some mechanism to switch between which pair of dimensions is being used at that time. Most examples used in this thesis use the drag-threshold mechanism of switching between dimensions, with Google Sketchup giving additional visual feedback through showing along which axis we are moving at each time by displaying text which proclaims that we are moving along the “blue” axis (the axes in Google Sketchup are identified primarily by colour). This however gives no predictive information, it does not suggest what we need to do to continue moving along the blue axis, or what we need to do to move along the red axis instead.

Some programs such as 3dStudioMax and AutoCAD have in the past used a four panel visualisation method, see Figure 15 on the previous page, where
each plane (which contains a pair of axes) receives its own window, and a fourth window displays the three dimensional object. This solution breaks all interaction into two dimensional components and thus negates one part of the problem (the 2d to 3d translation of movement) but instead moves the focus over to two other problems. These are: 1. Efficiency of movement and 2. Mentally projecting three dimensional movement from two dimensional movement. The first is simple to explain. Fitts’ law revolves around the distance a controller must move and target size, and shows that long movements of the controller are costly in time and precision as well as small targets being harder to hit (Benyon et al., 2005). Each of the three 2 dimensional panels offer movement in two of the dimensions, and the cursor must be moved between them to work with any axis which is not on offer in the panel on which the cursor is currently located. This results in frequent and long movement which equates to inefficiency. The other problem is that the user must now be able to mentally translate the movement which is needed in the three dimensional window into separate two dimensional components, and perform those actions on one object (one of the two dimensional representations) and follow the movement of the simulated three dimensional object in a separate window. Following multiple windows like this requires mental effort, and in fact, general usability guidelines for suggest that any manipulation of a control or object should always be done on the object itself, not on a separate controller (Cooper et al., 2007). The two dimensional windows have in effect become remote controllers for the three dimensional window. This results in most newer programs not using this solution, and indeed newer versions of programs such as AutoCAD and 3dStudioMax do not use this, even if older versions did use this four panel interaction. The four panel solution does solve the cognitive part of the problem as well as to simplify a certain part of the interaction, although Cooper et al. (2007) note that this comes at a significant cognitive cost of having to consolidate multiple viewpoints in the user’s mind to understand the location and movement of an object, as well as the longer movements, as previously mentioned, and difficulties with finding a particular object in three different viewpoints when the user wants to interact with the object. However, Cooper et al. (2007) do mention that this four panel method does allow for highly precise movement along an axis.

The four panel interaction and visualisation method illustrates one important point, and that is the two dimensional nature of the interaction being performed. In the four panel interaction method all interaction with the simulated three dimensional object is clearly and visibly collapsed into sets of two dimensional interactions. When examining the drag threshold method of interaction more closely it becomes apparent that the same is done there, but instead of visualising the collapse into two dimensions it is mostly hidden until after an action is performed. The non-visible nature of the possible actions seems to play a large role in the problems seen in the analysis. Although the currently used drag-threshold model of interaction constrains movement to pairs of two dimensional movement or even one dimensional movement, it does not display what the user must do to continue that movement or switch to moving along a different axis. This makes heavy demands on the user’s mental model of the situation, his internalised
understanding of the system, and requires that he vigilantly monitors the rotation of the environment outside of the object which is being manipulated. For this to be done in a reliable manner with full control requires high levels of internalised knowledge, which is only available to an expert in that particular CAD system, as different CAD systems using the same paradigm for interaction may still use slightly different settings to decide when a drag-threshold has been exceeded, and thus switch to another axis of motion.

Some design guidelines emerge from the analysis results. Here follow five guidelines for designing three dimensional interfaces for applications such as CAD programs. These guidelines are comprised of specific guidelines which result directly from the research which was performed in this thesis, as well as taking into account previously understood interface design guidelines which also apply for this domain. The guidelines which were developed based on this thesis aim to solve the problem of using a two dimensional tool to manipulate objects in a three dimensional virtual environment, and are based on the results which indicated that the central problem of interaction was the lack of clarity surrounding the shift in dimensions between layer 2 on the one hand and layers 3 and 4 on the other. These guidelines are:

1. Collapse the three dimensional interaction into multiple two dimensional movement options (matching the interaction method to the interaction tool),
2. keep those movement options visible (Cooper et al, 2007),
3. give a clear indicator as to how they relate to the possible movement directions of the cursor which is limited to two dimensions,
4. keep all movement options close to each other spatially to lower the need for movement and increase efficiency of movement according to Fitts' law (Benyon et al., 2007),
5. display those movement options on the simulated three dimensional object itself so that movement of the tool (cursor) and the resulting movement of the simulated three dimensional object are always visible at the same time.

This last suggestion does not come directly from the analysis, but rather from general user interface guidelines for performing actions on the item which is to be manipulated, not on a separate controller (Cooper et al., 2007).

The first guideline is based on the findings of this thesis and has been explored somewhat previously. Even so, it bears pointing out that anytime a two dimensional controller is used on a virtual three dimensional environment or object there must be some form of collapsing of dimensions, some sort of switching between the movement axes. Formalising that into a guideline is done so that this part of the design is done knowingly and with an understanding of what the designer is doing, making this design part explicit. The second design guideline is also based on what has been found in this thesis, and was landed upon in large part due to the focus on modern CAD systems such as AutoCAD and Google Sketchup which do not show this. A major difficulty in using these systems is that the user cannot see how a change in motion of the tool will affect the object being manipulated. This
guideline can be followed by multiple means such as by creating a visualisation in layer 3 (see figure 2, pg. 9) or by using the older four panel design shown in figure 15 on page 40, which brings us to the third and fourth guidelines. The third guideline is really a continuation of the second, and points out the necessity of thinking through any visualisation method so that it is simple and clear to understand how a movement of the mouse will translate into a movement of the simulated object being acted upon. The fourth guideline is then based on two distinct elements. One is that having multiple interaction areas which are spaced far apart is inefficient, which is shown by Fitts’ law (Cooper et al., 2007). The other is that working in one area while keeping attention on another and possibly having to mentally rotate and translate movements instead of easily seeing them on-screen is hard and should be avoided (Cooper et al., 2007). That is then continued in the fifth and last guideline which is based on previous guidelines on interface design and can be found in Cooper et al. (2007) and suggests that all manipulation should be done directly on the object being manipulated, and not on a separate control unit placed somewhere else.

An example of an interface designed using these guidelines is shown in figure 16. The interface would be shown in layer 3, the three dimensional interface layer (see figure 2, pg. 9), overlaid on the object to be manipulated. This interface consists of three intersecting and translucent circles which represent each of the three planes. Each circle follows the same rules of as the virtual world in which it resides, which results in the circles showing their orientation through the scaling effect of perspective. Each circle then has a slice extending out of its edge, somewhat like a pizza slice that is larger than the pizza. This slice further reinforces the perspective (scaling effect) and allows standard two dimensional controls (arrows) to be added for movement. As each slice is connected to a representation of that plane, these standard movement controls suggest movement along that plane. At the end of each slice there is a bent arrow which suggests rotation. This bent arrow allows for rotation on
that plane. This interface allows for collapsing all interaction down to one dimensional movement or pairs of two dimensional movement (guideline 1), keeps all movement options visible (guideline 2), displays which direction to move the cursor to effect movement (guideline 3), keeps all controls close to one another (guideline 4), and shows everything overlaid on the object itself (guideline 5). Figure 16 is of course a rough sketch of an interface, and shows only one of many possible implementations of these guidelines.

Even though the analysis did not fully answer the question posed in the beginning, which asked how existing research can explain how traditional two dimensional interaction tools allow for reaching into a three dimensional virtual world, and lacks a complete explanation of the cognitive processes involved it did clearly show where the problems are (the jump from two dimensional motion to three dimensional motion, the jump from layer 2 to layers 3 and 4), and how those problems can be bypassed, as is shown by the older 4 panel setup. That is an important finding even if the understanding of the the cognitive processing of the scenario is incomplete as it still results in possible solutions which can impact the usability and efficiency of programs in the domain being examined, as it allows the formation of guidelines to avoid the problem.

4.1. Discussion

The methodology used has had its shares of problems. I stand by my choice of a literature review to begin exploring this problem area within cognition and interaction design, and the decision not to attempt any experiments at this point seems to have been prudent. However, the analysis would have benefitted from the use of multiple scenarios, each representative of a popular method of making three dimensional interaction with a two dimensional controller possible. Multiple analyses would have been preferable using separate scenarios using the drag-threshold interaction method, a four panel interaction method with three two dimensional panels, an interaction method which would use layer 3 controls or indicators overlaid on the object in layer 4, as well as an interaction method using a three dimensional controller. As the literature on this interaction problem is sparse and exists in several different disciplines, each of which has its own focus and terminology, this was not obvious at the beginning of the research. The basic reason for not having analysed multiple interaction methods was thus mostly a lack of understanding of the problems of the domain, in particular that the necessary collapse of dimensions exists in all the multiple interaction methods. If this line of research is continued, then I would suggest a quick and efficient scenario/persona based analyses of each of those scenarios to help with designing any experiments around this interaction problem. Another difficulty was that the selected analysis methods also turned out to be complicated to use, and no clear guidelines were found to use these methods in this manner. A theory that claims to be useful for explaining actions would be easier to use for analysis if it includes a framework for using when performing analyses of actions, or even just suggestions or examples of how they can be used, and I would attempt finding other theories that do have usable analysis frameworks if this thesis were to be repeated. Having such a framework available would
allow for a more structured analysis using the theory to which the framework belongs, as well as possibly allowing adaptation of that framework to other theories which the researcher wishes to use.

As this is a literature review no ethical issues were encountered. Additionally, further research on this matter is unlikely to encounter any ethical issues, as no damage is likely to test subjects and no deception needs to be used for most forms of experiments to test this interaction and cognition problem. Some experiments might require deception of a nature which is unlikely to negatively affect the subjects in any way even without a de-briefing, although a de-briefing should of course be done if deception is used, if only so that the subjects can leave the experiment content with their own contribution.

A few things seem clear: There is a problem in understanding how a mouse cursor interacts with a three dimensional environment and objects, that problem has real world implications, and providing an explanation for that problem would be of immense help for designers and HCI researchers and would likely contribute to lowered training costs and lower error rates in the use of software such as CAD software. This lack of understanding becomes apparent when searching through both research on 3d interaction and when consulting books on user interface design, as was mentioned in chapter 1. This thesis has shed some light on the location of the problem and attempts to understand the nature of the problem, as well as identifying methods that can assist in circumventing this specific problem, all of which is a starting point for developing a more theoretical solution.

One of the most difficult challenges encountered turned out to be defining the problem clearly enough. Multiple books on design, as well as research on virtual environments or on children’s learning all turned out to ignore the source of the problem, focusing instead on the visible symptoms of the attempted solutions to the problem. Cooper et al., (2007) for instance mention the picking and threshold problems, but make no attempt to explain them and no mention of any elements making up the problem. Developmental psychology and other learning research explore the complexity of connecting tools to objects (Adolph et al., 1993) but even those dealing with virtual objects (Geuss et al., 2010) do not view the virtual domain as being different to the real domain, and treat the mouse-cursor as being equivalent to a hand reaching towards a real object. This is clearly not the case, or can in any case not be assumed without being backed up in some way. Adding to that, much of the work on affordance is purely theoretical in nature (Gaver, 1991; Dohn, 2009), and lacks guidelines on how theory can be put into practice. This all leads to a complete lack of material even exploring the complexity of linking a two dimensional tool to a three dimensional action in a virtual environment, which is an action done on a daily basis by whole professions such as engineers and architects. This lack of material exploring this problem is the reason why most of this thesis ends up revolving around clearly defining the problem, and showing real world consequences of the problem.

Some way of consolidating theories, or adding to an existing theory seems necessary to make them useful for real world problems that may not conform to scenarios which have been previously used. One such possible modification would be to the theory of affordance, and suggests that if actions could be said
to have and transmit affordances, as opposed to only the objects themselves providing information, then that would go far towards bridging a gap between perception-action and affordance, and give affordance a model which better allows for integration with theories of learning. That idea would then state that the information about what actions are made possible by an object can come from the object itself and/or from an action performed on that object, either be the actor himself or by a third party. This idea should work with Gibson’s (1986) original definition because of Gibson’s inclusion of movement and his focus on sampling sufficient information to get all the affordance information. Seeing Gibson’s mention of movement to sample information one is left to wonder whether there is really any difference between moving around an object to discover its affordances, or moving the object itself. Both fit well into perception-action loop theory, and so the question becomes whether the two may not be considered largely complementary when getting information from the environment. Is the perception-action loop a part of the discovery and transmission mechanism for affordances? This is only hypothetical, and will not be developed further in this thesis, but may provide an avenue for further work on consolidating theories of perception and getting them to work in domains and scenarios for which they were not originally designed. If this work were to be continued then it may be worth taking into account that Oliver (2005) suggests that the term affordance be discarded due to the confusion over definitions, but also suggests that any parts of the concept that can be shown to work should be kept and consolidated under a new term which would be more strictly defined from the beginning. As was shown in the background section the mouse cursor lacks a direct link to the object that needs manipulating but a temporary and invisible link can be created through pressing the mouse button. If seeing that action performed (or accidentally performing part of the action) transmits the affordances that were missing then the cursor has effectively gained an affordance after the relation between mouse and cursor was revealed. An example of an action giving information could be moving a closed container such as a milk carton, with the resulting movement giving information as to the amount of content, as well as what that container actually contains (is it milk, cottage cheese, or an equivalent weight in lead?). The movement gives information about the viscosity and level of substance in the container, the shifts in centre of mass even give information about the specific density of the contents. This then tells the user true information about the actions possible with that container, or affordances. This is not done through only visual perception in this case, but rather physical feedback, and an important aspect here is that a snapshot of the action will not give any of this useful information; it is the dynamic movement itself which gives information.

Returning to the research and the scenario which was explored and analysed we see that it is a very specific circumstance, and would be rare or non-existent except for the recent development of advanced three dimensional computer environments and the use of readily available interaction methods which are designed for working with a two dimensional desktop metaphor, but not for three dimensional work. It is even likely that an interaction problem of this nature has not existed before in the history of humankind. Perspective in drawing may be a related problem, as it requires an
understanding of how three dimensions can be simulated on a two-dimensional sheet by using a tool which is also constrained in two dimensions, but as that does not allow for interacting with the drawn object and moving it in the simulated three dimensions after the creation of the object then even that does not share this same interaction problem. Perspective drawing does not change the functionality of the tools during the drawing process; a smaller object drawn closer to the perspective point will always appear further away than a larger version of the same object drawn further away from the perspective point. The interaction problem might disappear or change drastically with the development of other display technologies and interaction methods, but may also always exist in some form so long as the interaction methods are not perfect simulations of the environment that is being manipulated. The mouse-cursor–virtual environment paradigm might thus be the clearest displays of this problem that we are likely to find, and therefore warrants our full attention as gaining an understanding the problem is likely to lead to large gains in efficiency in CAD programs as well as explaining interaction with other applications such as computer games where the use of a non-matching number of dimensions in the environment and the interaction tool can cause problems and create questions such as Rambusch’s (2010) question from page 1 about what tells the player how to reach into the environment and perform an action on the image of a door.

This thesis ends up showing how breaks in dimensionality between the tool and the environment causes problems, and suggests that the cause of the problems is the lack of a body-schema reference frame after limiting the reference frame to two dimensions at the start of the action. This aspect needs to be explored further and must be included in texts on interface designs so that future interface designers can understand how interaction must be broken down into dimensions that match the tool being used as well as the importance of showing the interaction possibilities on the screen. This simple explanation goes far to support the practical goal of this thesis: to provide tools to UI designers to support the use of readily available interaction methods in three dimensional computer systems.
References


