



UPPSALA
UNIVERSITET

IT Licentiate theses
2016-006

Context

The Abstract Term for the Concrete

ANTON AXELSSON

UPPSALA UNIVERSITY
Department of Information Technology





UPPSALA
UNIVERSITET

Context

Anton Axelsson
anton.axelsson@it.uu.se

April 2016

*Division of Visual Information and Interaction
Department of Information Technology
Uppsala University
Box 337
SE-751 05 Uppsala
Sweden*

<http://www.it.uu.se/>

Dissertation for the degree of Licentiate of Philosophy in Computer Science with
specialisation in Human-Computer Interaction

© Anton Axelsson 2016
ISSN 1404-5117

Image on page vii courtesy of FCIT
Printed by the Department of Information Technology, Uppsala University, Sweden

Abstract

This thesis deals with the term ‘context’ and the aim has been to reason about the term in order to see whether it is possible to reach a satisfactory understanding of the concept. But the thesis is also a journey into human reasoning and conveys a certain view of human cognition. It aims to synthesise results of studies within psychology, cognitive science, anthropology, and human-computer interaction. My understanding is that context is not something we are a part of, but rather something we create mentally in relation a specific goal. Determination of something ambiguous thus comes from top-down processes related to a goal. I believe context has been wrongly interpreted in HCI as that which a user is situated *in* and which a product is being used *in*. I suggest instead a separation between the user environment and the user context.

List of Papers

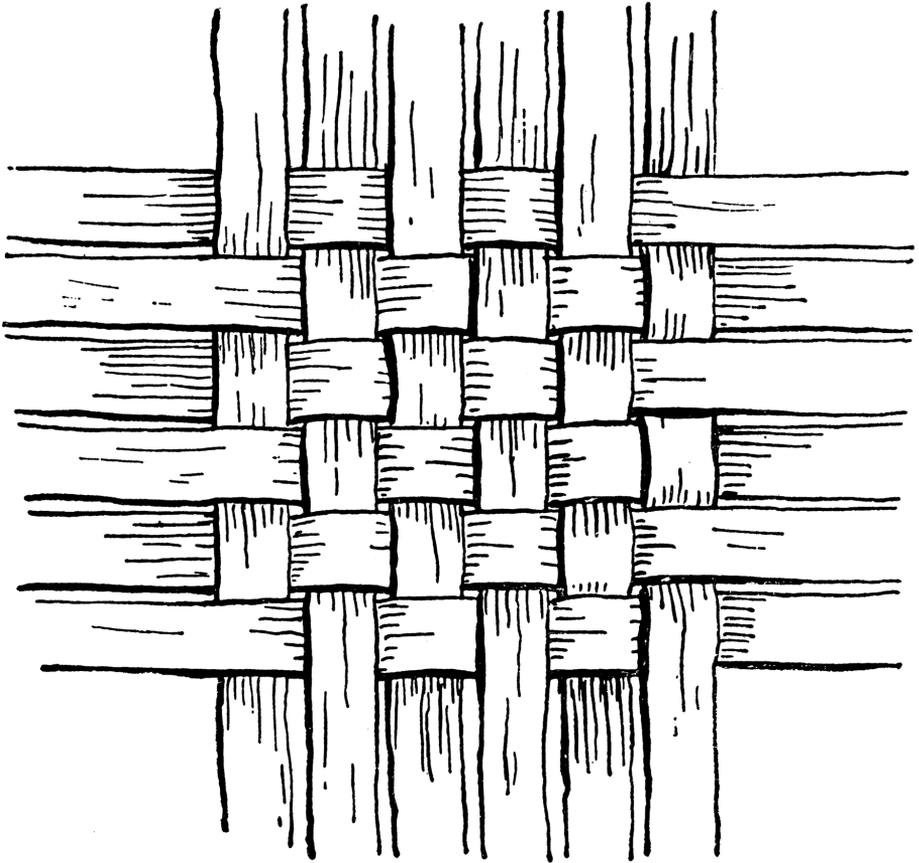
This thesis is based on the following papers

- I** Axelsson, A., Andersson, R., & Gulz, A. (2016). Scaffolding executive function capabilities via play-&-learn software for preschoolers. *Journal of Educational Psychology, 108*. Advance online publication. <http://dx.doi.org/10.1037/edu0000099>
- II** Jansson, A., Erlandsson, M., & Axelsson, A. (2015). Collegial Verbalisation — the value of an independent observer: an ecological approach. *Theoretical Issues in Ergonomics Science, 16*, 474–494. <http://dx.doi.org/10.1080/1463922X.2015.1027322>
- III** Bodin, I., Axelsson, A., Vännström, J., & Jansson, A. (Submitted). Eliciting strategies for evolutionary in-vehicle design: exploring the hypothesis of context-free strategy categories.

Other Publications

The works listed below are publications by the author which are not included in the present thesis.

- Krupenia, S., Jansson, A., Fagerlönn, J., Bodin, I., Ahlm, V., Larsson, S., Vännström, J., Axelsson, A., & Frimalm, R. (Submitted). Developing a 1st Iteration Concept HMI for Supervising and Controlling a Self Driving Truck.
- Krupenia, S., Jansson, A., Fagerlönn, J., Grane, C., Vännström, J., Bodin, I., Axelsson, A., Bromander, E., Ahlm, V., Nicola, D., Larsson, S., Frimalm, R., & Friström, J. (Submitted). Developing and Assessing a 2nd and 3rd Iteration Concept HMI for Supervising and Controlling a Self-Driving Truck.
- Haake, M., Axelsson, A., Clausen-Bruun, & M., Gulz, A. (2015). Scaffolding mentalizing via a play-&-learn game for preschoolers. *Computers and education*, 90: 13–23. <http://dx.doi.org/10.1016/j.compedu.2015.09.003>
- Jansson, A., Stensson, P., Bodin, I., Axelsson, & A., Tschirner, S. (2014). Authority and level of automation: Lessons to be learned in design of in-vehicle assistance systems. In *Masaaki Kurosu (red.) Human-Computer Interaction*, Springer Berlin/Heidelberg. 413–424. http://dx.doi.org/10.1007/978-3-319-07227-2_39
- Axelsson, A., Anderberg, E., Haake, M. (2013). Can preschoolers profit from a teachable agent based play-and-learn game in mathematics?. In *H. C. Lane, K. Yacef, J. Mostow & P. Pavlik (red.) Artificial Intelligence in Education*, Springer Berlin/Heidelberg. 289–298. http://dx.doi.org/10.1007/978-3-642-39112-5_30
- Anderberg, E., Axelsson, A., Bengtsson, S., Håakansson, M. & Lindberg, L. (2013). Exploring the use of a teachable agent in a mathematical computer game for preschoolers. *Lund University Cognitive Studies*, 154 , 161–171.



*The word 'context' originates in the Latin verb 'contextere'
which means 'to weave'*

Prologue

“How can we confront the blooming, buzzing confusion that is ‘context’ and still produce generalizable research results?”

Nardi (1995, p. 70)

This thesis deals with the term ‘context’ and the aim has been to reason about the term in order to see whether it is possible to reach a satisfactory understanding of the concept. But the thesis is also a journey into human reasoning and conveys a certain view of human cognition. It aims to synthesise results of studies within psychology, cognitive science, anthropology, and human-computer interaction along with philosophical views underpinning human reasoning as well as ontological and epistemological foundations. My experience with other, similar works within information technology is that they usually either focus on users *or* developers and thus creates an illusion that there would be an underlying difference between user and developer cognition. The aim here, is to marry these two views in the sense that what is true for a user’s cognition is also true for a developer. In early years of human development, many neural connections are created and throughout life many of these connections are lost and others are strengthened which means that individually we formulate somewhat different brains by specialising in the environments in which we individually thrive. However, the mechanisms by which this happens is no different from whether you are a developer of systems or a user of the same systems. Any biases you as an experimental researcher or system developer discover in users will also be true for yourself and I believe this to be important to understand in order to bridge the epistemological gap between developers and users. I urge you as a reader to have this in mind when reading the present thesis.

Anton Axelsson
Uppsala, 2016

Contents

1 Disclaimer	1
Purpose	2
Underpinnings	2
Reality for Me and the Reality in Itself	4
The Regularity Assumption	4
The Environmental Interface Assumption	5
Reality, Environment, and Cognition: The Russian Doll of Experience	6
Distinction of Terms	6
2 Context to Date	9
Definitions in HCI	10
Other Sources	12
Everything That Surrounds?	13
Aspects of Context	15
Conclusion	16
3 The Deep Blue Sea	17
Situated Cognition	17
Embodied	18
Embedded	19
Extended	20
Situated Research in the Lab and in the Field	20
Memory	20
Problem-Solving and Decision Making	21
Learning and Knowledge Transfer	22
Scaffolding Mature Cognitive Behaviour (Paper I)	25
Learning by Teaching	25
Teachable Agent	25
Cognitive Growth	26
The Preschool Study	26

Conclusion	28
4 Touching Reality	31
Methods for Eliciting Context	31
Verbalisation Protocols	32
Collegial Verbalisation (Paper II)	33
5 The Narrow Straits	37
Reasoning in the Lab	37
Wason Selection Task	38
Providing Context	39
Developing for a Future World (Paper III)	41
Strategies Analysis	41
Formative Approach	42
Conclusion	42
6 Discussion	45
Who Provides the Context?	46
Goal-Dependence	47
Foundation of Attention	48
Reifying Context for the Reader	50
Conclusion	51

Chapter 1

Disclaimer

“With an undefined notion of context, as with an indefinite future, anything is possible.”

Clark (1996, p. 92)

The purpose of this introductory chapter is to provide a theoretical framework from which this thesis can be interpreted and to give my epistemological and ontological view. In academia, we tend to aim at finding ‘truths’ which are, unfortunately, rarely of any help when faced with real-life problems. This can be handled by creating nicely trimmed models and definitions which can be considered pragmatic tools with the effect of compromising ‘the truth’. To me, within science and theoretical analysis, we need consider Socrates’ motto that the only thing one can truly know is that one knows nothing. This, I interpret, as that there is always a more correct answer out there and we can never truly know if we have reached ‘the truth’, however, there are answers that are more or less useful. This is the foundation of the pragmatic theory of truth (e.g., Peirce, 1905) which I hold very dearly. With this disclaimer I emphasise that what is described in this thesis is a simplification of reality with the goal of enhancing designers’ understanding, who do not work theoretically, of human cognition with the honest footnote that it is purposely simplified to offer a pragmatic value in daily work of interaction designers.

One example which most target readers should be familiar with where compromise has been employed is the ISO 9241-11 definition of usability:

Extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use. (ISO, 1998, p. 2)

This definition was developed with the purpose of being used when evaluating developed products (Bevan & Macleod, 1994). The definition has been heavily criticised for being out-of-touch and not capturing the complexity that is reality. Personally I find it useful as a model which enlightens important, though not all-embracing, aspects to consider in evaluation of a product.¹ The definition is easy to grasp and apply since we as designers understand the concept of a user, a goal, and what effectiveness, efficiency, and satisfaction is. The definition does however contain an elusive term which meaning escapes most of us: ‘context’. I would like to throw light on this term and try to make it more graspable. This I intend to do by exposing the complexity in launching systems in the wild and anchoring my view of context in research of situated action. This thesis also contains three papers from my own research. These are very disparate and are included to throw light on ‘context’ from different perspectives.

Purpose

The aim with this particular thesis is to try and gain understanding of the term; to build a more theoretical base for what context could be considered to be. There are many uses of the term and it seems to be *something*, but I have doubt we really understand what it is, and we are thus not able to make proper use of it as a term. We have to empirically test our tools in order to reveal whether they are of any use to us or not. The term context is frequently used and now we need to start evaluating whether it is a usable concept or not. If it turns out not to be, we have learnt something valuable, but if we can make use of it we have gained something extraordinary: a tool, i.e., the embodiment of a theory of how an entity can be used to reach certain goals. This thesis concerns an empirical and theoretical endeavour into showing the value of the term context as an analytical tool in Human-Computer Interaction (henceforth, HCI).

Underpinnings

In this endeavour, I will argue for that context is not so much something which we are in but rather something we formulate when trying to act in the world. Context seems to be something which a problem-solver formulates; a sort of mental representation or mental environment in which solutions can be allowed to emerge and outcomes can be simulated and theorised.

¹I also find it useful when thinking about design although I am perfectly aware that it was not created for that specific purpose.

To grasp this view it is necessary to consider the distinction between reality and our perception of reality which has epistemological implications. Throughout history of science and philosophy, there has been a constant battle between mind and matter. In metaphysics there has been the idealist versus the materialist, in philosophy of mind there has been the dualist versus the monist, in philosophy of science there has been the rationalist versus the empiricist, in philosophy of truth there has been the coherence theorist versus the correspondence theorist, in psychology there has been the structuralist versus the behaviourist, and, in cognitive science there has been the cognitivist versus the connectionist. However, between these extremes there has in the end always emerged a synthesis of mind and matter: functionalism, pragmatism, ecological psychology, and now gaining popularity in cognitive science: Situated Cognition. What these synthesised perspectives tell us is that we need *both*; mind *and* matter, organism *and* environment, correspondence (*with* nature) *and* coherence (*in* nature).

Fundamental to my view of knowledge is the dichotomous separation between *knowing* and *understanding*. There are other ways of naming the parts of this dichotomy, in memory research it is often phrased as implicit (know-how, procedural, or tacit) and explicit (know-that or declarative) knowledge. Implicit knowledge is that which we can perform practically such as driving a car, juggling, or eating thus usually coupled with motor and procedural memory. This is a form of tacit knowledge that is not easily verbalised. Explicit knowledge on the other hand is more related to our semantic knowledge of facts, hypotheses, and theories and is that which we can talk about².

This is, for me, a dualist view based in the separation of acting in an environment (reality or body) and experiencing reality (mentality or mind). Knowing is something very situated and dependent on environmental factors. Our bodies are well equipped at automating motor actions in relation to incoming perceptual stimuli (i.e., perception-action cycle or sensorimotor abilities). To me, this view makes it clear what the difference between morals and ethics are, morals is what we practice and we have a 'gut-feeling' of what is right and wrong when we are in specific situations, whilst ethics is the theoretical understanding of what we should do in general. Thus knowing is very practical and tangible and understanding is very theoretical and an abstraction of knowledge where specific, situational factors are kept at a minimum.

²I personally do not like the idea of implicit knowledge because there is nothing particularly implicit about juggling, for example, it is something I can very much make explicit by showing that I know how to do it. If a distinction should be made to emphasise the ability to talk about knowledge or not it is better to make the distinction between verbalisable and non-verbalisable knowledge.

To me, language is a very practical type of knowledge and relates very much to knowing. There is today very little problem for us to use the term ‘context’ because we know what it is. However, I sense that there is little *understanding* of the term. Once we have a more abstracted view of a concept we can make better use of it, after all, theories are tools that we can use and the better theories we develop, the more possibility there will be for knowledge transfer between different problem domains.

Reality for Me and the Reality in Itself

The German philosopher Immanuel Kant pointed out that human knowledge is limited in its scope. The reason, he believed, is based in his dichotomy of *phenomenal* and *noumenal* reality (Kant, 1922/1781). This distinction lies in the sensual and non-sensual duality of objects. When we perceive an object, we do not experience it as it truly exists, instead we perceive it through our limited senses. For example, when you look at an object, what forms the image of the object are the photons reflected onto the object, you do not actually see the object in itself. Thus in our perceptions, we only experience the phenomenon of the object and not the intelligible object; not the noumenon.

I will hijack Kant’s dichotomy in my attempt at grasping the impalpable notion of context, doing so with the appreciation that I will not do his ideas justice (to some readers it will most likely be perceived as out of context). However, the distinction between what is *real* and what is *experienced as real*, which thus reveals the limits of our understanding, is fundamental in how I view human cognition and therefore also how I interpret the concept of ‘context’.

The Regularity Assumption

I will make no attempt to describe what noumena might be, it could be that they are, in their essence, the results of quantum mechanics as hypothesised by physicists, or we might simply be part of an elaborate computer simulation. This is of no interest to me and definitely outside the scope of the present thesis. I do, however, take for granted that reality is made up of regularities which we can detect cognitively by acting through our bodily perception-action processes in the engagement with reality. Reality for me is simply the realm of the noumena that make up our world, it is the true nature in itself which we cannot fully comprehend due to the limits of our senses.

The Environmental Interface Assumption

Important to understand is that when I use the term ‘environment’ I specifically mean the part of reality in which an agent is situated and has the ability to perceive. Although we cannot perceive the noumena of reality directly, I contend that we can interact with noumena using our body (e.g., by kicking or lifting objects). This is because our body is part of reality and therefore a noumenon in itself and it therefore affects reality. Your body is not just a vessel moving through noumenal reality, it is an integral part of it. An organism cannot be understood without the aspects of reality which has evolutionary shaped it, physically and therefore also cognitively.

I think it is common to misconceive the environment as only the things we perceive with our bodies and thus separate our body from the environment (environment starts where my fingertips end). I would therefore like to make a discrimination similar to Kant’s between reality and environment. I see environment as that which we perceive our bodies to be within in a given moment and it is bounded by a perceptual horizon. The environment is, if you like, the interface in which we engage with reality. We communicate our environments by using domain terms such as ‘forest’, ‘kitchen’, ‘office’, ‘garden’ and so forth, a domain thus being a generalised version of similar environments. An environment is thus the closest enclosure we consider ourselves to be in and is limited by what we can perceive when situated in that domain. My view of environment overlaps with Jakob von Uexküll’s *umwelt* (i.e., a self-centred world; Kull, 2010). A more generalised term such as ‘forest’, I consider to be a domain and the environment would thus be a particular forest which an organism is situated in.³ So, I may not be able to perceive noumena in all their essences, but interact with them I surly can. And although we do not perceive reality as it is, we do perceive some resulting effects of interaction with reality through our perceptions which helps us in drawing mental inferences about reality.

In as much as I am a monist and a materialist, I cannot shake my fundamental experience of a separation between mind and the material. This dualism is, however, based in properties, not in Cartesian substances. I assume that mind arises from the same matter as the basis of perceived phenomenal objects. But, I do believe in the importance of separating mental experiences from physical experiences. As previously stated, I consider reality to be constituted in regularities and therefore a deterministic system composed of an unfathomable amount of variables (or dimensions). Our senses, however, only allows us to perceive reality in a limited number of

³If you are familiar with object oriented programming, the distinction can be made between environment as the object instance and the domain as the class, or, from a biology perspective it is the distinction between an individual and its specie.

dimensions (spatially only 3). This makes our interaction with reality problematic and thus the environment seems probabilistic or even stochastic to us. Our minds are, in a sense, statistical inference machines trying to handle this discrepancy (for an interesting overview of the mind as operating through predictive inference, see Hohwy, 2013).

This view of environment might be difficult to grasp, because it goes somewhat against the common conception of environment. Viewing environment as an interface to reality is a first-person perspective rather than an outsider perspective. What I see as environment is simply those parts of reality that have been made perceivable by an organism in order to facilitate its survival.

Reality, Environment, and Cognition: The Russian Doll of Experience

The relation between reality, environment, and cognition is illustrated in Figure 1.1. What this illustration attempts to show is that the environment is the stage on which the phenomenal experience of reality through our senses and the direct interaction with reality through our motor abilities take place. With the environment we engage in real-time, in an online mode (see On-Line and Off-Line Cognition in Chapter 3). This view suggests that an environment is a perceptual representation of a limited part of reality, whilst memories in the mind are constituted by concepts in long-term memory which has been consolidated in previous experiences within different domains.

Distinction of Terms

I refrain from using the term context as something in which an individual is situated, but rather use it as an analytical term when problem-solvers are trying to mentally represent particular environments. In the present thesis I will therefore avoid using the term ‘context’ when talking about the environment or environmental factors. It is common practice to use the term context when referring to environmental factors. I will instead use the term ‘environmental specificity’ when discrepancies between two different environments need be considered. That is, separate parts of the reality of one individual. For one observer-specific part of reality across individuals I will instead use the term ‘domain’ which would be a generalisation of several environments. An environment is thus a limited part of reality, perceptually and cognitively available to an individual organism.

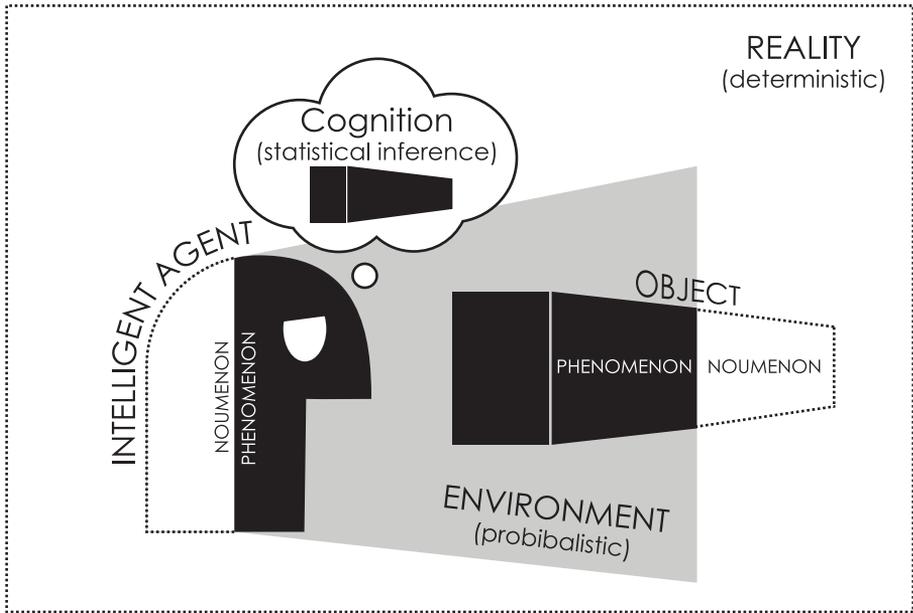


Figure 1.1: Caricature of the relation between cognition, environment, and reality

Now that this theoretical view has been laid on the table, we will start looking at how the term ‘context’ has been defined previously. And in the succeeding chapters, we will look at the trade-off problem between rigour (the control of variables) and relevance (generalisability). I will try to argue that environmental specificities lies at the heart of this problem and that context is a dynamic model we build in order to try and weed out what specific environmental factors disrupts performance results in implementation of, for example, experiments and deployed systems.

Chapter 2

Context to Date

“That one’s ability to retrieve (or recognize) an item is heavily influenced by the relation between that item’s storage and retrieval contexts is indisputable. ‘Context,’ however, is a kind of conceptual garbage can ...”

Smith, Glenberg, and Bjork (1978, p. 342)

The amount of research related to context is overwhelming. It is an important factor in language science, history, and arts. There are numerous reviews, overviews, and surveys on the research of context (e.g., Akman, 2002; Brézillon, 1996; Chalmers, 2004; Clark & Carlson, 1981; Wan, 2009). There has also been several special issues on the matter,¹ and for the past 20 years there has even been a biennial conference concerning any research where context is crucial.²

In the early 20th century, Harvey Carr was one of the first to conduct research on what is seen as contextual influences within psychology (1917; 1925; cited in Smith et al., 1978). Thus, the term has morphed from being only related to words and their semantics to being relevant for physical objects and their influences on participants’ performance in experiments. From here, it has then been picked up and used within any situated research and eventually turned up within HCI.

¹Some examples related to HCI are: *Human-Computer Interaction* (16:2), *Information Processing & Management* (38:5), and *Science, Technology, & Human values* (37:4)

²<http://context17.org/>

Definitions in HCI

Definitions of the term ‘context’ are sparse throughout HCI literature. In early literature on ubiquitous and context-aware computing, definitions of context are found wanting. The term was simply used in agreement with colloquial interpretation. As the field matured, the need to understand the notion of context grew. Early attempts consisted in either listing aspects considered relevant for context or providing synonyms. Dey (2001) pointed to the difficulty with definition by choice of synonyms or listing of examples in applying these to his own work: “When we want to determine whether a type of information not listed in the definition is context or not, it is not clear how we can use the definition to solve the dilemma” (p. 2). Initially, positioning in the form of GPS was more or less the only environmental aspect utilised. Schmidt, Beigl, and Gellersen (1999) acknowledged that there is more to context than location and they presented mobile devices that would adjust screen brightness and orientation based on surrounding lighting conditions and device rotation, now universal features of today’s smartphones.

As technology turned evermore mobile in the 1990’s, the need to build flexible and context-sensitive devices grew. With it, understanding of environmental impact on usability, and thus defining context, became a necessity. Attempts were made, and still are, at building models and taxonomies of context for pervasive or ubiquitous computing (e.g., Chen, Finin, & Joshi, 2003; Henriksen, Indulska, & Rakotonirainy, 2002; Reignier, Brdiczka, Vaufreydaz, Crowley, & Maisonnasse, 2007; Rodden, Cheverst, Davies, & Dix, 1998; Schmidt et al., 1999; Winograd, 2001).

A collection of descriptions of the term ‘context’ found in HCI literature is listed in Table 2.1. One oft-cited definition of context within HCI and pervasive computing listed in the table is the one provided by Dey and Abowd (2000). This definition is an elaborated version of the definitions provided by Schilit and Theimer (1994) and Pascoe (1998). The definition has been criticised for having too much of an immediate environmental focus. Chalmers (2004) pointed out that:

Such definitions are common in context-aware and ubiquitous computing, but they do tend to emphasise objective features that can be tracked and recorded relatively easily, and to de-emphasise or avoid aspects of the user experience such as subjectively perceived features and the way that past experience of similar contexts may influence current activity—issues. (p. 223)

Table 2.1: A collection of elaborations on, and definitions of, the term ‘context’ from HCI and ubiquitous computing research

Definition	Citation
Context-aware computing is the ability of a mobile user’s applications to discover and react to changes in the environment they are situated in. In our system mobile users run software that is constantly monitoring, or <i>subscribing</i> to, information about the world around them.	Schilit and Theimer (1994, original emphasis)
Three important aspects of context are: where you are, who you are with, and what resources are nearby. [...] Context encompasses more than just the user’s location, because other things of interest are also mobile and changing. Context includes lighting, noise level, network connectivity, communication costs, communication bandwidth, and even the social situation: e.g., whether you are with your manager or with a co-worker.	Schilit, Adams, and Want (1994)
Users, tasks, equipment (hardware, software and materials), and the physical and social environments in which a product is used. [...] The context of use consists of those components of the work system which are treated as given when specifying or measuring usability.	ISO (1998, context of use)
Context is a subjective concept that is defined by the entity that perceives it. For example, one entity may conceive of its context as location whereas another may view it from a temporal perspective. It could also be a more ethereal construct, e.g. the emotional state of a person. Therefore, context could be generally described as the subset of physical and conceptual states of interest to a particular entity.	Pascoe (1998)

- A context describes a situation and the environment a device or user is in Schmidt et al. (1999, model for context)
- A context is identified by a unique name
- For each context a set of features is relevant
- For each relevant feature a range of values is determined (implicit or explicit) by the context

Context is any information that can be used to characterize the situation of an entity. An entity is a person place or object that is considered relevant to the interaction between a user and an application including the user and applications themselves.

Dey and Abowd (2000)

[A] context can be seen as a network of situations defined in a common state space.

Crowley, Coutaz, Rey, and Reigner (2002)

Context can mean many things; it might be the tasks that the system is being used to perform, the reasons for which the tasks are being carried out, the settings within which the work is conducted, or other factors that surround the user and the system. The context, though, is as much social as technical.

Dourish (2004)

[C]ontext refers to information about a person's proximate environment such as location and identities of nearby people and objects. [...] we consider activity as a type of contextual information which can be used to characterize the situation of a person

Loke (2004)

Other Sources

The definitions from HCI leave a lot to wish for. They tend to include pretty much ever aspect of reality and that is not very helpful when trying to understand a term so we must look elsewhere. Scientific definitions of the term 'context' are difficult to come by. There is, as with any concept or term within science, little consensus on how it should be defined. Definitions have ranged from consisting of external (e.g., "stimulation from the external environment", McGroch, 1939, p. 347) to internal (e.g., "inner slates of the

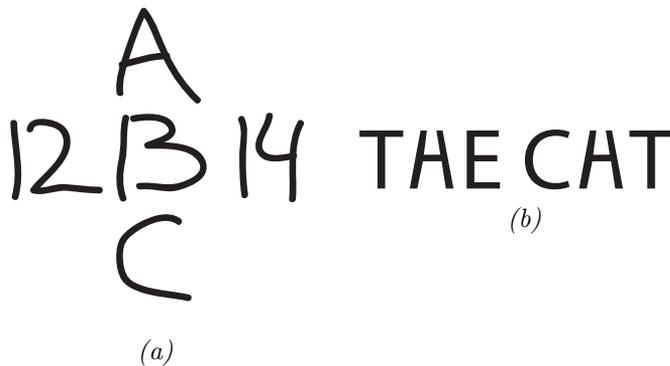


Figure 2.1: Examples of ambiguous symbols

experiencing person”, Reiff & Scheerer, 1959, p. 19) factors in relation to participants of psychological studies. Clark and Carlson (1981) argue that the term has been extended to the point that its denotation has become murky. Perhaps this is why the term has been considered a ‘conceptual garbage can’ as quoted in this chapter’s epigraph.

The usual way of exemplifying context is by showing how ambiguous symbols are interpreted differently depending on what other symbols surround them. Two examples are shown in Figure 2.1. The first variant (Figure 2.1a) shows how a symbol can represent either the letter ‘B’ when symbols are read vertically or as the number ‘13’ when symbols are read horizontally. Similarly, in the second variant (Figure 2.1b) we see the term ‘THE CAT’ and the A-symbol is considered an ‘H’ in the word ‘THE’ and an ‘A’ in the word ‘CAT’.

In the first monolingual dictionary of the English language, Cawdrey (1604) defined the term as “the agreeing of the matter going before, with that which followeth”. The term was originally only considered relevant in relation to discourse and Clark and Carlson (1981) explains that the standard definition of context is “parts of a discourse that surround a word of [sic] passage and can throw light upon its meaning” (p. 314).

Everything That Surrounds?

When you read words, you do it holistically. Your visual sense takes in a lot of data which your perceptual system must process to handle ambiguity. I suspect that this process is dependent on your past experiences, in that, the more you encounter something the more likely you are to quickly process that something mentally due to the statistical inferential abilities of the brain. Consider the phrase ‘TAE CAT’. You have probably encountered the phrase ‘THE CAT’ more often than ‘TAE CAT’, ‘THE CHT’, or ‘TAE

A
12 B 14
C

Figure 2.2: Example of unambiguous symbols

CHT’. However, if you were a cat lover and there existed a race of cats called Tae then you possibly interpret the phrase as ‘TAE CAT’, or if you spend much of your time working with an organisation called CHT then you are arguably more likely to be able to interpret the phrase as ‘THE CHT’. This suggests that when encountering the ambiguous phrase ‘TAE CAT’, you are interacting with the environment and previous experience with the environment determine how you handle the ambiguity. However, when you realise that what represents the ‘H’ and the ‘A’, in what you thought to be spelling out ‘THE CAT’, you notice that they look exactly the same (i.e., ‘A’). Therefore you start analysing how this is possible and that is when you put the A-symbol in focus and you draw the conclusion that it is what surrounds the symbol that determines how it is interpreted. You are thus creating a context in which you are analysing a phenomenon, and try to answer a specific question: ‘How can I interpret the same symbol in two ways?’.

To me, a more specific and telling answer to this question would be that the symbol simply looks like an ‘A’ *and* an ‘H’. If there existed a letter ‘A’ in your language, then that is what you probably would have perceived. Similarly, if you would have used an ‘H’ instead of the A-symbol, the phrase would have been interpreted as being misspelled. Now we can, more or less, perfectly read the phrase because the A-symbol looks enough like an ‘A’ and enough like an ‘H’ to handle the situation. Consider Figure 2.2, in this case ‘context’ — when defined as that which surrounds something — is not very important when interpreting the middle letter as a ‘B’.

To further illustrate my point, when you read something like: ‘Waht wuold yuo lkie fro diennr tnogiht?’ you once again have little trouble determining what the intended message is. This is called *typoglycemia* and our ability to interpret the text comes from the fact that there is enough redundant information to interpret the intended message. Grainger and Whitney (2004) found that words with transpositioned letters is easier to read than words where letters have been substituted. Thus, a word’s specific letters

are crucial for identification and comprehension and “readers cannot rely exclusively on context for word recognition” (Rayner, White, Johnson, & Liversedge, 2006, p. 193).

Furthermore, it seems reasonable to argue that this ability results from the fact that the intended words are very common in the English vocabulary. More uncommon words would probably be a lot more difficult to understand when letters are jumbled. Attributing the handling of ambiguities to the colloquial concept of context, at least to me, adds little explanatory value.

Aspects of Context

Thus, HCI definitions of ‘context’ are too broad, engulfing everything in its path whilst the standard definition seems too vague with little explanatory value. According to Wan (2009), the idea of context was introduced in artificial intelligence (AI) by Weyhrauch (1980) and in this field it is formalised through propositional or predicate logic. Turner (1993) identified five crucial aspects of context in relation to AI: (1) context recognition facilitates prediction and biases situation assessment in an intelligent agent, (2) context modulates behaviour by facilitating goal formulation, (3) context focuses an intelligent agent’s attention, (4) context makes appropriate actions salient for reaching a particular goal, and (5) context facilitates handling contingencies.

Clark and Carlson (1981) identified 6 aspects of context from reviewing uses of the term: (1) context is information, (2) it is subjective, (3) it relates to a process being carried out, (4) it relates to a specific occasion, (5) context is only constituted by information currently available, and (6) the information which is considered context must be able to interact with the ongoing process. This led them to define ‘context’ as “information that is available to a particular person for interaction with a particular process on a particular occasion” (p. 318). The authors also make a distinction between *intrinsic* and *incidental* context where the former consists of information necessary to investigate a particular psychological process, and the latter is aspects that indirectly affects process performance such as hunger, mood, aches, or fatigue or information present in stimuli which are irrelevant to the process.

These aspects seems to capture more relevant aspects of context than any other I have come across. They seem to suggest that context has more to do with what an individual perceives than what is actually out there in reality which HCI definitions seem to centre on. Baars (1988) argue that context is nothing but a mental phenomenon and it is an unconscious dito.

“Contexts include currently unconscious expectations that shape

conscious experiences, and currently unconscious intentions that shape voluntary actions [...] the physical environment affects our experiences and actions only if it is represented in the inner world. Thus the context-in-the-world inevitably shapes our experience by way of the ‘context-in-the-head’.” (p. 110).

Conclusion

It is a tremendous paradox that a term which is meant to represent the concrete and specific can be so abstract and general. To understand what this concept really means we need to take a deeper look at human cognition by looking at how humans function in situated reality where context seems to have most impact. This is what is in store for the next chapter. In succeeding chapters we will then also take a look at what happens in situations less rich in context before getting in to a final discussion of this elusive term.

Chapter 3

The Deep Blue Sea

“... context is not so much something into which someone is put, but an order of behavior of which one is a part.”

McDermott (1993, p. 290)

As designers we often make the assumption that users will see connections and draw parallels to other usage situations when engaging with our developed products, culminating in an intuitive interaction. The desktop metaphor (Smith, Irby, Kimball, & Harslem, 1982) is one of the more successful metaphors that introduced white-collars to computers by designing the user interface to resemble common artefacts found in the office (folders, files, printers, in-boxes etc.). These concise metaphors are difficult to find and even when found they might not intuitively signify the designer’s intended functionality for the actual user. It is also the case that when you develop expertise in a certain domain, you have the ability to think a lot more abstract and draw parallels. Few users are tech-savvy and therefore think more concretely about the specific usage situation, thus knowledge transfer from interaction with physical objects into the virtual world might not occur. Usability is very dependent on situational and environmental factors and to clarify how, we will now take a look at research carried out which outlines the complexity of being out in reality; out on the deep blue sea.

Situated Cognition

It has been argued by Hollan, Hutchins, and Kirsh (2000) that situated cognition should be considered the theoretical underpinning of HCI. According to Robins and Aydede (2008), the research paradigm of situated cognition is

based on three basic assumptions: (1) cognition is *embodied* and is therefore dependent on mind *and* body to function, (2) actions based in cognition is making use of the environment of which the mind is *embedded*, and (3) cognition *extends* out into the environment and beyond the boundaries of the body. These are three subareas which have an increasing dependency on environment from embodiedness, through embeddedness, to extendedness.

Embodied

The idea that the mind is embodied is grounded in that cognition is dependent on what can be experienced through sensorimotor capacities of the body and, further, that each sensorimotor capacity is in itself dependent on environmental, cognitive, and bodily influences (Varela, Thompson, & Rosch, 1991). A ground breaking idea of the mind as embodied came from George Lakoff when he realised that language is very much built up on metaphors, and the body very often acts as the referent of these metaphors (Lakoff & Johnson, 1980). For example, we talk about the *foot* of a hill or the *head* of a beer. We also use words that express positions which are based in the body. When we are ill or hurt, for example, we lie down or when we are sad our whole body droops which makes us think about negative feelings as being *down* and the opposite, when we are happy and well we hold our posture straight and might be skipping which makes us think about positive feelings as being *up*.

This relation is a two-way street where bodily actions influences emotions as much as emotions influences the body. In a social psychology study, English speakers were given Chinese ideographs whilst either (1) pushing up on a table from below, (2) pushing down on a table, (3) or not pushing at all (Cacioppo, Priester, & Berntson, 1993). Participants who were pushing up on the table rated the ideographs more positive than those participants who either pushed down or did no pushing at all.

Even the mere conception of body influences cognitive behaviour; personality can change with the conception of your own body or a character you assume. Cognitive capabilities or social attitudes can change and be aligned with biased views of stereotypes when you yourself fit that stereotype. For example, using immersive virtual reality, researchers have been able to give participants changed bodily representations by making them either look older or give participants a changed ethnic appearance. What consistently is found in these experiments is a reduction of implicit biases against the outgroups being virtually represented (for an overview see Maister, Slater, Sanchez-Vives, & Tsakiris, 2015). A less elaborate, but just as telling example comes from an experiment within educational technology. Students were asked to solve mathematical problems on a computer and

were randomly assigned either a male or a female digital avatar. What this study found was that students assigned to the male avatar characters performed better at solving the mathematical problems than students who were assigned female avatars, regardless of the student's own gender. The embodied gender stereotype of women being less able in mathematics influenced the participants' performance (Lee, Nass, & Bailenson, 2014).

On-Line and Off-Line Cognition

Within situated cognition and embodiment, a distinction between on-line and off-line cognition is made. The Chinese ideograph study is a perfect example of on-line cognition where motor *behaviour* shapes attitudes. This can be contrasted with off-line cognition where motor areas of the brain are used in imagining moving parts of the body without *actually* moving them. What this suggests is that both perceptual and motor functions of the brain which evolved to engage with the environment, can be resourced internally in thinking and problem-solving. According to Wilson (2002), "the function of these sensorimotor resources is to run a simulation of some aspect of the physical world, as a means of representing information or drawing inferences".

Embedded

To understand how a brain function, it is not enough to just consider the individual. Herbert Simon has analogy of an ant on a beach which illustrates this. If we only study the path an ant has taken by analysing the traces in the sand, the behaviour of the ant looks very complicated. However, if we also take into account the obstacles along the path such as sticks and stones we realise that much of the behaviour is orchestrated by the structure of the environment.

Viewed as a geometric figure, the ant's path is irregular, complex, hard to describe. But its complexity is really a complexity in the surface of the beach, not a complexity in the ant. (Simon, 1996, p. 51)

Humans tend to off-load cognition to the environment. For example, Kirsh and Maglio (1994) noted that people playing Tetris seemed to involve themselves in seemingly unnecessary actions. Players tended to continuously rotate the Tetris zoids which is an action which does not get you directly closer to your goal of placing the zoid at the bottom of the screen. However, time to mentally rotate objects has been shown to be linearly increasing with the number of degrees of rotation (Shephard & Metzler, 1971). Mentally

rotating zoids takes more time than the motor action of pressing a key and letting the computer rotate the zoid for you. This turns out to be a very effective strategy for quickly identifying spots where the zoids would fit. Thus Kirsh and Maglio made a distinction between pragmatic (goal-oriented) and *epistemic* (cognition-relieving) actions.

Extended

It is not just that cognition is happening as a result of being in a body or embedded in an environment. Clark and Chalmers (1998) argue further that technology can be seen as an extension of mind. This suggests that calculators and calendars is a way to off-load cognitive abilities into the environment, or cognitive abilities can also be distributed between individuals. Hutchins (1995) describe how a pilot's task of flying a plane makes the cockpit a whole cognitive system. This extended view of cognition relates to the idea of distributed cognition (Hutchins & Klausen, 2000). Hutchins sees a cockpit as a cohesive cognitive system and argues that cognition is not only inside the pilot's head, but distributed within a sociotechnical system.

Situated Research in the Lab and in the Field

Any researcher that has conducted social, psychological, or cognitive studies will appreciate the trade-off between study rigour and relevance of results. There are always variables uncounted for that influences participants in ways we as experimenters cannot foresee. This is what we generally call extraneous or confounding factors, that is, the discrepancy between our abstracted view of the study object and its manifestation in a situated reality.

Whenever an unexpected result occurs, as researchers I think we often attribute these confounding factors to context. It certainly holds true for HCI where implementation of a new, working system might not yield expected results when the intended users starts interacting with it. This is usually a result from discrepancies in technical knowledge between designer and user and/or discrepancies in domain knowledge between user and designer. We will return to these issues of discrepancy in Chapter 4. As background knowledge, studies related to environmental specificity can throw light on these discrepancy issues.

Memory

In a now classical study of human memory, Godden and Baddeley (1975) let participants learn a set of words being either on land or under water. Half of the participants on land then went into the water whilst half of

the participants in the water were called up on land. After the participants recital of the memorised list, it was revealed that memory recall was better in the environment in which encoding had taken place, regardless of whether it was on land or in water. This suggests that retrieval of memories is easier if situated in a similar environment to that in which encoding of these memories were performed.

Problem-Solving and Decision Making

Studies of problem-solving and decision making have mostly been conducted in laboratories using students as subjects. However, in the 1980's Gary Klein set out to study decision making in the field by observing professional firefighters throughout their workday. Within the classical decision making research paradigm, the approach has been to focus on normative decision making, that is, studying how decisions can be made optimally. What this entails is that the decision process is a way of determining what to choose from a set of options by weighing pros and cons of each option and choose the option with the highest utility. Klein and his associates instead focused on a descriptive approach which has been termed Naturalistic Decision Making. In contrast to normative research they describe how decisions are actually made rather than how decisions should be made. Through the classical paradigm humans have time and time again been shown to be quite bad at making decisions in the lab. When Klein went out into the field of firefighters he saw that these people were excellent problem-solvers. A surprising result of the research was that firefighters do not seem to make decisions by considering multiple options and choosing the optimal solution. Instead, they tend to quickly grasp the situation and intuitively come out with a plan which is mentally simulated or quickly discussed and then implemented (Klein, Calderwood, & Clinton-Cirocco, 2010; Okoli, Weller, & Watt, 2016). Making decisions in a lab means you often have plenty of time contemplating options. Firefighting is very different, it is time-limited and you have a lot at stake; you have to act quickly in the situation in order to save lives. This is a rich and real situation which the firefighters are thrown into, as opposed to lab experiments where decisions are based on arbitrary information with low-stake consequences for uninformed decisions. Often, the goal of the task is not clear to the participant of a lab experiment of decision making and contextual information is sparse. But, when you are out on the deep blue sea, and you are a trained professional with years of experience, then intuitions kick in aided by the specificities of the environment and time for optimal solutions do not exist, instead, action is key.

Recognition-Primed Decision Model

Through the firefight study, Klein and associates created the Recognition-Primed Decision (RPD) model (Klein, 1993). The model is depicted in Figure 3.1. The aim of this model is to fill a gap in the decision research since normative approaches do not handle decision making in time-pressured and ill-defined tasks. In the model, the decision maker's experience is utilised in mental simulations which is a vital part in solving problems quickly without deliberating on multiple options, and very often implementing the first formulated action plan. Thus the decision maker has to *satisfice* (Simon, 1956) since the dynamicity and time pressure of the situation bounds the decision makers rationality. This means that the decision maker favours a satisfactory and sufficient, but not necessarily optimal, solution. The course of action is determined by formulating plausible goals through the use of relevant cues in the environment along with expectations grounded in similarly experienced situations.

Learning and Knowledge Transfer

In studies of classical conditioning, environmental specificity tend to be an unwanted side-effect impeding on study rigour. When conditioned by series of clicks in relation to food delivery in one surrounding, rats tend to be less likely to approach the site of food delivery when the conditioned stimulus is presented in another surrounding (Hall & Honey, 1989). This side-effect can be attenuated through hippocampal lesions. The hippocampus is an area which is part of the limbic system of the medial temporal lobe in the brain of vertebrates. This area plays a crucial role in the formation of new memories about experienced events, and therefore detection of novel stimuli. Many hippocampal neurons have so called place fields which means that certain neurons fire whenever its host passes through a particular environment and is thus crucial for spatial memory and navigation. A comparison between London taxi and bus drivers showed that taxi drivers tend to have more grey matter in the mid-posterior hippocampi than bus drivers (Maguire, Woollett, & Spiers, 2006). This is attributed to the fact that taxi drivers must learn to navigate a city independent of a predefined route.

When conditioned to fear by shock in one chamber with both different and similar stimuli to a chamber in which shocks are not administered, rodents with the hippocampus removed have shown to elicit fear responses in both environments. Thus they are still conditioned to the shock but cannot discriminate between the two environments (Frankland, Cestari, Filipkowski, McDonald, & Silva, 1998; Wiltgen, Sanders, Anagnostaras, Sage, & Fanselow, 2006). This comes to show that anything present in this kind

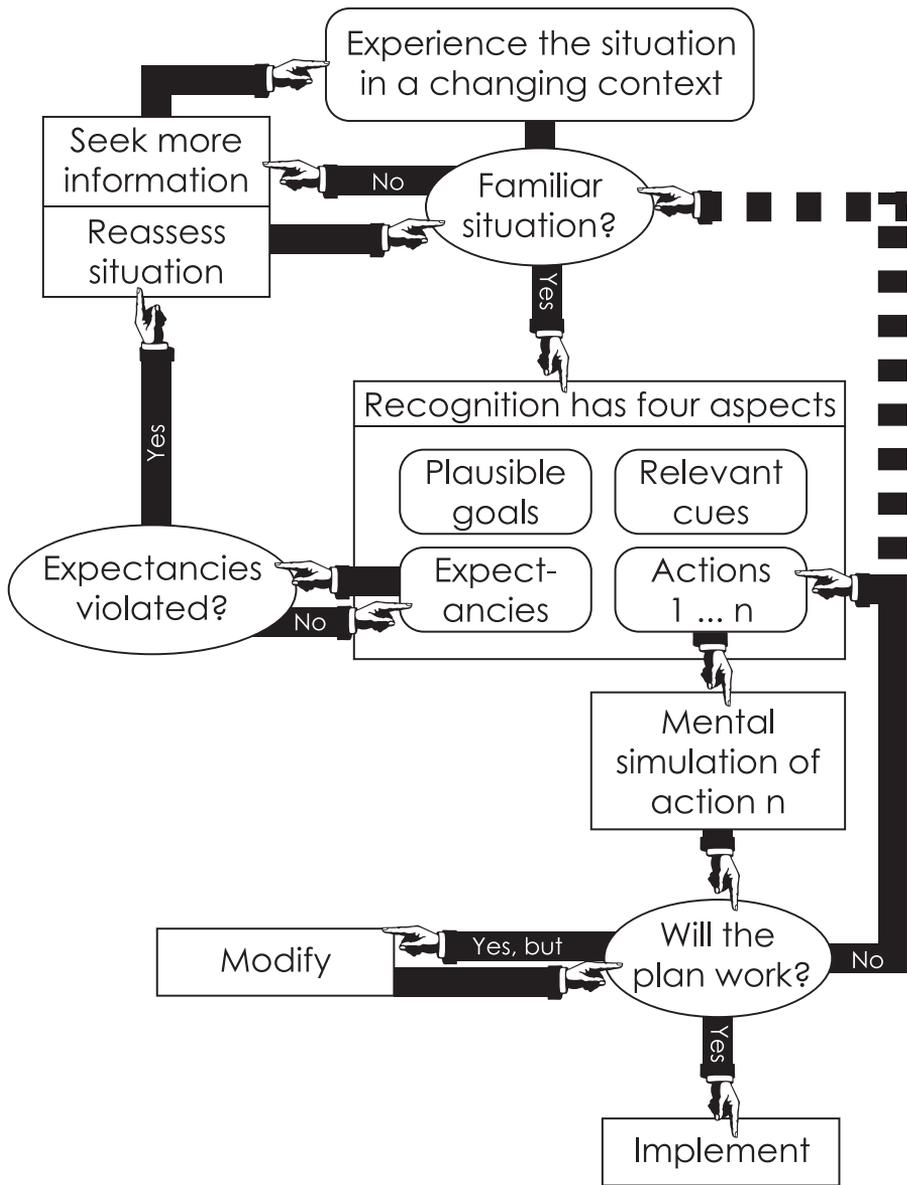


Figure 3.1: The Recognition Primed Decision model redrawn from Klein (1993)

of experimental set-up should be considered a stimulus. However, experimenters would generally only consider the intended and manipulated stimuli as important and thus the experimenter formulates an experiment context which is in essence different from the subject's environment and this would account for the discrepancy between expected and actual research results.

Similar problems have been encountered when investigating the difference between formal and informal education. Anyone familiar with abstracted forms of arithmetic should be able to do simple calculations about anything. However, if you are taught arithmetic informally, then the objects which you perform calculations with are of utmost importance. Gay and Cole (1967) investigated formal and informal education amongst the Kpelle people of Liberia. In the Kpelle language, numbers are represented as "one of a thing, two of a thing", and so forth. When these people are put in school and subjected to formal education, they have problems understanding mathematics because for them arithmetic manipulations are equivalent to manipulation of physical objects but in western formal education, actual physical objects are rarely part of the equation, instead numbers are the objects in themselves (Scribner & Cole, 1973). This poses problems which is difficult for a western teacher to grasp because the pupils do have arithmetic skills but they cannot be transferred to the classroom setting where manipulation of numbers serves no obvious purpose. This could be likened with the lab rats that do not elicit a conditioned response although vital stimuli are present because these stimuli have no meaning in the particular setting. Studies on knowledge transfer show time and time again that we humans are notoriously bad at generalising our skills across domains. Numerous teacher anecdotes and anthropological studies testify to that some people can solve arithmetic problems in everyday situations such as transforming recipes, calculating bowling scores, or handling currency when buying or selling goods, but when asked to perform similar arithmetical tasks in a classroom, these people fail (Chaiklin & Lave, 1993; Gay & Cole, 1967; Lave, 1988).

Vygotsky (1986) made a distinction between everyday and scientific concepts which conceptualises the difference between formal and informal learning. Everyday concepts are formulated bottom-up by a learner being exposed time and time again to different instances of the concept through everyday interactions. These concepts become well-known but difficult to define. In formal education, on the other hand, scientific concepts are taught and formalised top-down in the sense that the learner is given a term and its definition and must then try to understand aspects of reality that relates to the term. People with an informal education background placed in a formal setting has been observed not to ask 'why' questions as frequently as people with a formal education background (Fortes, 1938 cited in Scribner & Cole, 1973), the alleged reason being that children taught informally will

have their education in the “real world” where, similar to the fireground, the purpose of doing something is intrinsic to the situation.

What these disparate studies show is the impact of environmental specificities, from how we are influenced by environmental stimuli helping us in cognitive tasks such as remembering, to how lack of contextual factors can render us helpless in problem-solving situations. We will now look in more detail on a study conducted with preschoolers playing an educational game and see how environmental factors can seemingly scaffold mature cognitive behaviour in young children.

Scaffolding Mature Cognitive Behaviour (Paper I)

Something which is important to understand when developing products is the fact that the product itself will also be an integrated part of the user environment. The way we choose to develop a system and its interface will have a strong impact on the user’s interpretation of possible actions and thus it affects user’s cognition in ways exemplified by the discussed studies above. I was involved in a project with the Educational Technology Group at the Lund University Cognitive Science Department where we developed a computer game around preschool mathematics (Anderberg, Axelsson, Bengtsson, Håkansson, & Lindberg, 2013).

Learning by Teaching

Our mathematics game uses a digital implementation of the learning by teaching (LBT) paradigm presented in the seminal work by Bargh and Schul (1980). The paradigm has proven fruitful with regard to cognitive benefits for learners. Instead of learning for oneself, students are instructed to tutor their peers which increases students’ motivation to learn. This motivation arises through social aspects where the tutoring students take responsibility for their peers’ learning (i.e., the protégé effect; Chase, Chin, Oppezzo, & Schwartz, 2009) and thus have to know the study material well. Furthermore, the paradigm seems to stimulate meta-cognitive reflection where students starts thinking about learning and problem-solving processes (Flavell, 1979; Schwartz et al., 2009).

Teachable Agent

LBT computer games uses a computerised tutee — called a *teachable agent* (TA; Blair, Schwartz, Biswas, & Leelawong, 2006) — which is in essence an artificial intelligence algorithm that can be taught by a player. This computer implemented version of LBT has been proven to have the same edu-

cational benefits as the traditional paradigm time and time again (Biswas, Leelawong, Schwartz, Vye, & The Teachable Agents Group at Vanderbilt, 2005; Gulz, Haake, & Silvervarg, 2011; Ogan et al., 2012; Pareto, Haake, Lindström, Sjöden, & Gulz, 2012). However, studies conducted so far have focused on primary school children aged between 8 and 14. With our game, we were the first to initiate research to investigate whether benefits of this paradigm could be witnessed also with preschoolers.

Cognitive Growth

It is well established that children of preschool age goes through a phase of vast cognitive changes in developing both social and communicative skills as well as executive functions (Clements & Perner, 1994; Hughes, 1998; Perner & Lang, 1999). In order to benefit from LBT, tutoring students must be able to keep focus on their tutee's actions and follow the tutee's progress in learning. This brings about doubt whether digital LBT implementations would be suitable for children as young as 3 to 6 years old. We decided to investigate preschoolers' ability to pay attention to a TA.

A pilot study was conducted at a local preschool (Axelsson, Anderberg, & Haake, 2013) with 9 participants and they seemed to enjoy playing this preschool mathematics game and they also seemed able to engage with the TA. Therefore we planned a larger study in which we brought the game and eye-tracking equipment to a preschool in a rural area in the south of Sweden.

The Preschool Study

The complete study involved a battery of pre-tests of mathematics and language skills, as well as theory of mind and executive functions (complete study details can be found in Paper I and in Haake, Axelsson, Clausen-Bruun, & Gulz, 2015). The main test consisted of the participants playing our LBT resembling game, and into the game visual distractions were incorporated in order to experimentally simulate a noisy environment. The distractions were implemented in the form of animations which were played at educationally crucial moments where the tutoring participant would need to attend to the TA. This was in a mode of the game where the TA is playing by himself and is choosing numbered buttons. In order to understand how well their tutee is doing, the tutoring participants would have to attend to the TA's actions. Animations were played just before the TA makes his choice and also after the TA has made his choice when it does not matter whether participants are attending the TA or not. We thus investigated whether participants would be able to inhibit the distractions in favour for attending to their digital tutee when the tutee displays his skills.



Figure 3.2: Heatmap during competing visual stimuli. Fixations are averaged across all participants. Fixated areas are colour coded along the gradient scale provided at the bottom of the figure, i.e., the more red an area is the more it was fixated by participants. Visual distractions were animated at the top part of the screen, in this case an aeroplane flying past.

One of the pre-tests was a standard inhibition test called an *anti-saccade task* (Hallett, 1978).¹ In line with previous studies (Fukushima, Hatta, & Fukushima, 2000; Munoz & Everling, 2004), the preschoolers who participated in this study were terrible at inhibiting in this standardised pre-test. The unique finding of this study was, however, that in the main test children paid much more attention to the TA than to the visual distractions (Figure 3.2) which indicates that the game creates a sort of cognitive scaffold which enables them to inhibit far better than in the pre-test. Of course the participants were attending to the visual distractions, but only after the TA had made his choice (Figure 3.3). During the TA's choice, the majority of the participants did not look at the distractions at all.

¹Eye movements between fixations are called 'saccades' and in the anti-saccade task participants usually focus on a cross in the middle of a computer screen and now and again a dot appears at some location on-screen. The participant's task is to saccade away from the cross in the opposite direction of where the dot appears (i.e., look away from the dot). Thus deliberately inhibiting the natural response of looking at the dot.

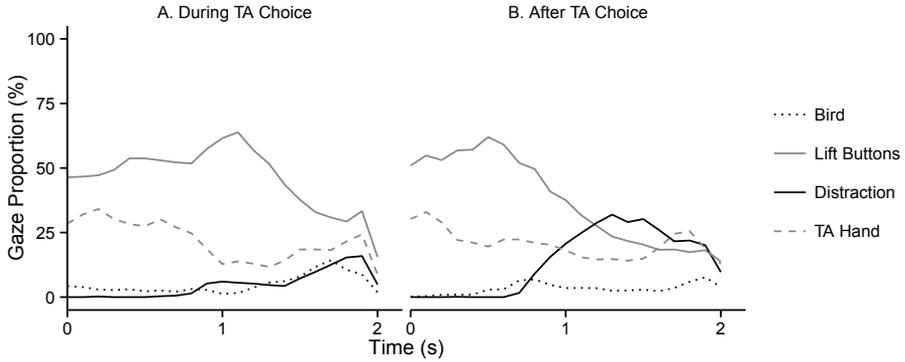


Figure 3.3: Gaze proportion of four areas of interest over time during (A) and after (B) TA choice. Time 0 on the x-axis denotes distraction onset.

This study implies that environmental specificity can act as a scaffold for eliciting more mature cognitive behaviour; in this case, in terms of executive function capabilities. There were also indications in the study that the LBT game also elicited more mature mentalising capabilities in the participating preschoolers (Haake et al., 2015).

Conclusion

One could consider that this chapter has been all about context in the colloquial sense and how it affects performance. In my view, what has been clear is that cognitive performances are definitely influenced by environmental factors as dictated by perceptual abilities and experiences. In our study of preschoolers playing an educational game, we concluded that the discrepancy in children’s ability to inhibit in the main task but not so much in the pre-test came down to context. Using my now more nuanced view of context, the results had very little to do with it. What we meant by the conclusion was that there were more intricate factors influencing the children in the main task than in the pre-test. Not knowing what these factors were we generalised it to context. In retrospect this is a somewhat unsatisfactory answer. The unique finding of this study is the fact that it shows the intricate relationship between different executive functions. The ability to stay focused is not just about developed volitional inhibition processes but also how attention is guided by motivational factors that assists in filtering out irrelevant sensory data. This, of course has to be specifically addressed in experiments tailored to investigate specific aspects of attention in order to really understand the intricate relations between these executive functions. Thus, the only thing we could conclude was that context — in the common

sense interpretation — mattered.

One implication of the findings discussed in the present chapter for HCI and systems design is the obvious need for testing newly developed systems before deployment. When the new system is left in the hands of its specified users, there are no guarantees that it will be functioning to the degree of satisfaction (if at all) as intended by the designer. The users technical knowledge, previous experiences, as well as surrounding environmental factors could all impede on the human-computer interaction. User interface elements that you as a developer have created in the belief that the design clearly signifies the elements functionality might not at all trigger as a signifier for the user. You cannot be sure that your chosen metaphors work. So whilst developing, it is thus crucial that you test your product with users. And before, as well as during, development you must try to elicit as much information as possible about the intended users and the environment into which the designed product is meant to fit. Similar to the firefighters, the more you are immersing yourself in the target environment, the more obvious the strategy for designing your product will become and development will consist more of intuitive problem-solving rather than deliberating over design choices.

Chapter 4

Touching Reality

“... the inner context preserves important information from the past, which is not available from our current surroundings at all. It makes more sense, therefore, to locate the psychological context inside the nervous system.”

Baars (1988, p. 110)

The fact that a user’s cognition is influenced by environmental factors is taken advantage of in elicitation methods. Human memory is highly associative and has a strong coupling with the environment, after all, memories are based in and used for interaction with reality. When developing systems for expert users, knowledge of the work domain is crucial which means that any decent researcher needs to get as close to actual target environments as possible.

Knowledge elicitation has two fundamental problems: (1) as designers we assume that users can verbalise their needs (the tacit problem) and (2) elicitation usually is carried out in a domain very different from the work environment (environmental cue problem), often in a conference room far away from where the product is meant to be used. To handle this, ethnographic methods are employed and researchers will have to come out of their ivory towers and get out into reality, close to the intended users.

Methods for Eliciting Context

In order to understand user environments there are methods developed to help designers form abstractions of these environments through which the products developed are meant to be interacted. Usability Context Analysis (Thomas & Bevan, 1996) is a method by which a design team brings a

number of stakeholders to what the creators of the method calls a Context Meeting. The result of this meeting is a document statement describing the product to be created, who its intended users are, the tasks the product is meant to support, and the circumstances of product use. By bringing together users, domain experts, designers and other stakeholders to this discussion, the method handles the first of the two fundamental problems of the elicitation process by collecting data from multiple sources. It does, however, not solve the second problem since the elicitation location is usually a conference room.

Another method is Contextual Inquiry (Beyer & Holtzblatt, 1998) which is a qualitative user-centred research method in which designers follow along side-by-side with intended product users throughout their daily work routines. Interviews are conducted whilst observing the user and the user and researcher collaborate to make it clear what the user is doing and why. This means that the researcher not only discusses things with the user but also observes what the user does. This handles the second fundamental problem of knowledge elicitation discussed above. The point here is to get very close to the intended users and understand how a future product could be fitted in their working life. However, the method does not fully handle the first problem. Of course the researcher has the ability to witness the user first hand which helps, but the researcher has often very limited knowledge of the work domain which can make it difficult to know what questions to ask and when in order to get further explications of user behaviour.

In order to handle both these fundamental problems, recruitment of domain experts can be of tremendous help. In our research team we have developed a verbalisation technique which handles this issue by letting colleagues verbalise on each others behaviour with data from the work place rather than relying only on interviews or only on operator verbalisations.

Verbalisation Protocols

Verbalisation methods are very common in knowledge elicitation within HCI. Most common is the concurrent verbalisation protocol where, for example, operators would perform work whilst explaining what they are doing (think-aloud). This has the limitation that adding the task of verbalising might interfere with the operators' work and thus not give a fair rendition of a normal working shift. Another limitation is that if the operators are immersed in a task with a heavy load on cognitive processes, the operators tend to stop verbalising and these difficult tasks are, paradoxically, often the most important to have explicated in an elicitation study. Another variant is the retrospective verbalisation protocol where operators would verbalise after

completing work, either by watching a video of themselves performing work or from sheer memory. The limitation with this protocol is that operators might not remember exactly what they were thinking and thus start confabulating, or they might not be honest in order to save face in case they remember that they made some mistakes.

The distinction between formal education or scientific concepts and informal education or everyday concepts discussed in the Section Learning and Knowledge Transfer of Chapter 3 can also have an impact on verbalisation results. Educated participants in psychological experiments often provide more telling verbal accounts of how they solve a task than less educated participants (Scribner & Cole, 1973). This means that the level of education of your verbaliser could affect both the quality and quantity of verbalisation data.

Collegial Verbalisation (Paper II)

To handle these limitations, our research group has developed a new form of verbalisation protocol where we let operators verbalise whilst watching videos of their colleagues performing work. This we have termed *conspective verbalisation* in order to distinguish it from other protocols. This could be seen as being concurrent verbalising since the operators are verbalising as events are unravelling. Or, it could be seen as retrospective verbalisation since operators are verbalising on something that has already happened. However, it is important to understand that this protocol is a fundamentally different form of verbalisation since the verbalising operator is another person than the performing operator. The verbaliser could essentially be a colleague or a domain expert. This protocol takes care of the issue of being emotionally invested in verbalising on one's own behaviour whilst sparing the verbaliser's cognitive resources.

This protocol has then been used in conjunction with either a concurrent or retrospective protocol to form a compound method called *Collegial Verbalisation*. Crucial to this method is using colleagues who share a work environment, and have done so for many years. It is very important that verbalisers are well familiar with the workplace environment because the environment is loaded with situated cues to memories shared between the operators. When observing colleagues perform work, verbalisers are able to access generalised memories of being in similar situations and thus explaining what might be going through the mind of their colleagues (Figure 4.1). These verbal reports can then be compared with verbalisation reports of the target operator and these data in conjunction creates a more objective view of the work tasks than any of the constituent verbalisation techniques do alone.

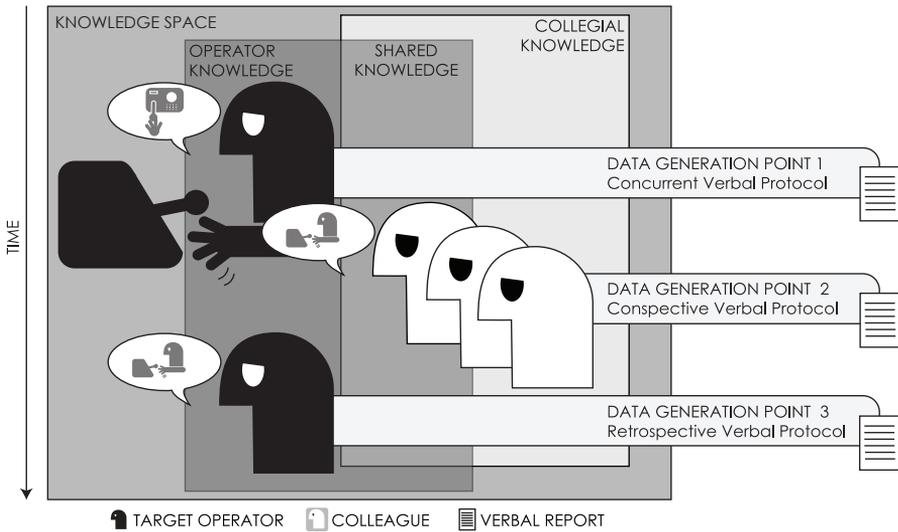


Figure 4.1: The compound Collegial Verbalisation method

The memory model behind this compound protocol is depicted in Figure 4.2. The idea is that operators are acting on stimuli in their work environment; this is recorded on video and serves as stimuli for later verbalisers. If concurrent verbal protocol is used, then the conspective verbalisation sessions are performed with muted recordings. All operators in the different verbalisation sessions are well familiar with the work environment. The whole workplace and worker interaction environment is viewed as a generalised memory in each operator. When a target operator is working, the specific situation is loaded to working memory and both consolidated to long-term memory and offloaded in the form of a verbal report, if concurrent verbal protocol is used. If retrospective protocol is used, it can be performed immediately after the work task has finished and then the specific memory of the situation will be more readily available to the verbaliser — loaded from long-term memory with the help of the video recording. The longer the gap between performance and verbalisation, the more generalised the memory of the specific situation will become since it will be integrated with earlier and later, equivalent experiences. With regard to the conspective protocol, only generalised memories of similar situations will be available to the verbaliser since this person has not actually experienced the specific situation first-hand. Instead, data in the form of educated speculations of a colleague's behaviour will be collected and compared between different verbalisers.

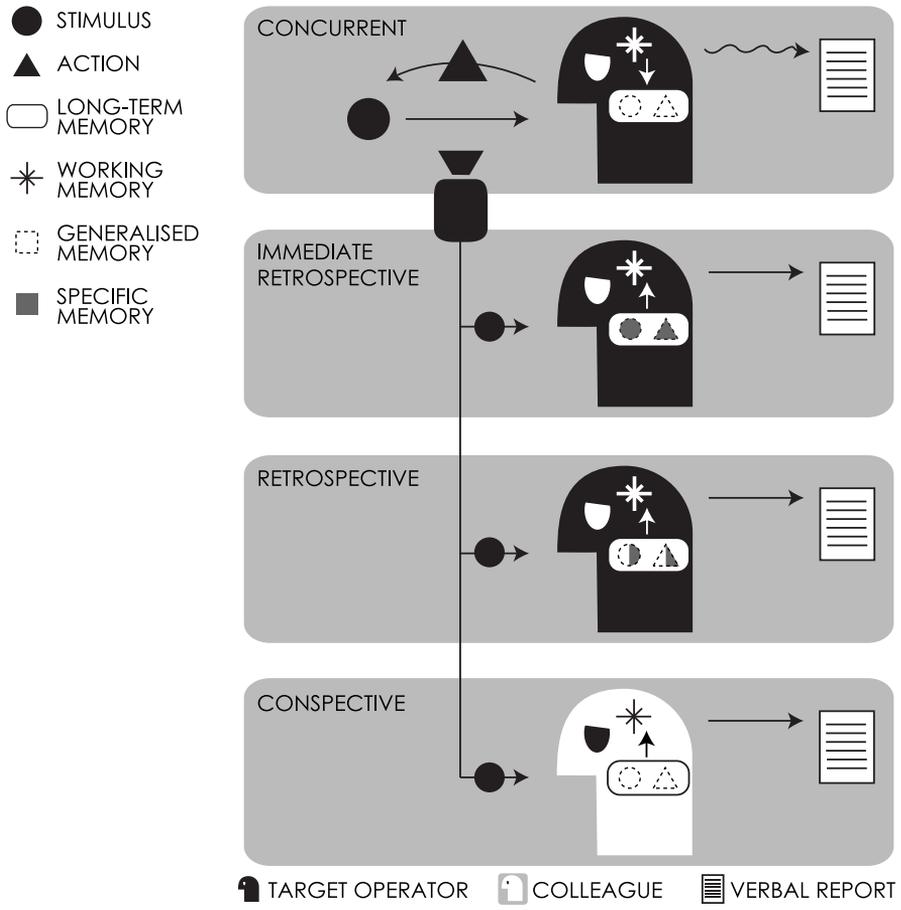


Figure 4.2: Joint memory based in experience of similar environmental interactions

Chapter 5

The Narrow Straits

“The focal event is placed on center stage, while context constitutes the stage itself.”

Goodwin and Duranti (1992, p. 9)

Great care must be taken when working in the narrow straits of our research ‘labs’. The understanding of a user’s domain will always be limited and when we are away from stakeholders as well as actual users and the target environment, it is easy to produce a system using short cuts and only implementing features that handle the most common situations and tasks a user encounter, leaving the system vulnerable to contingencies. To exemplify some problems, we will take a look at some research conducted on human reasoning in the lab. This gives an insight both to how humans reason but also to some strange effects that rarely occurs outside the lab and enlightens why system development can go horribly wrong when the developer has abstracted away too far from reality.

Reasoning in the Lab

There is a large body of research within judgement and decision making carried out in the lab. Researchers often conclude from the results of such experiments that human reasoning is flawed and tainted by numerous cognitive biases. A bias is a tendency to consistently reason in a certain way which contradicts rules of formal logic or statistical probability.

For example, Peter Wason discovered what he termed *confirmation bias* after he conducted an experiment with the so called 2-4-6 problem (1960). This is a rule discovery task where the participants’ task is to find an underlying rule for a given triad of numbers. First participants are given the triad

(e.g., 2, 4, and 6) and are asked to collect evidence of the rule by providing the experimenter with other sets of number triads. The experimenter will then verify or deny whether a given triad follows the experimenter's confabulated rule. When participants feel that they have collected enough evidence they then explicate the rule. Participants quickly find a pattern in the numbers but the problem is that they tend to only confirm their hypothesised rule by providing triads that follow this rule. This results in participants often giving the explanation that the underlying rule is three numbers where 2 is added to the previous number. However, the experimenter's rule is actually any three numbers in ascending order. The only way a participant can discover this rule is by providing a triad of numbers that do not comply with their own rule (e.g., 1, 5, and 7). This would still follow the experimenter's rule and thus the participants would gain valuable information in order to reject their hypothesised rule and formulate a new hypothesis.

Another famous example of cognitive bias was discovered by Tversky and Kahneman (1981) through numerous experiments on what they call the *conjunction fallacy*. The most famous version of this experiment is the so called Linda problem where a description of a fictitious person outlining how Linda majored in philosophy and was concerned with issues of discrimination and social injustice. The participants are then given a list of possible things Linda is involved in today, such as that she is a teacher, social worker, or that she is a member of the League of Women Voters. Along with these alternatives are also the option whether Linda is active in the feminist movement, and the option that she is a bank teller and active in the feminist movement. The participants' task is to rank these options in order of what is more probable. What participants tend to do is to rank Linda as a bank teller and active in the feminist movement higher than Linda being only active in the feminist movement. From a probability point of view, this result is somewhat surprising since it cannot be more probable that Linda would be *both* a bank teller and active in the feminist movement rather than just active in the feminist movement.

Some 100 more biases have been discovered over the years through decision-making, memory, and social psychology research, such as the *base rate fallacy*, *anchoring bias*, *IKEA effect*, and *loss aversion* to name a few.

Wason Selection Task

One of the most investigated tasks within problem-solving is Wason's Selection Task. It was originally conducted as a small study by Wason (1966). It turned out to be a quite difficult task for participants and the results of the task are still not fully understood. This reasoning task has been reproduced

and redesigned in a multitude of variations since, giving disparate and conflicting results on human reasoning. It is a task of formal, logical reasoning of the form: $P \rightarrow Q$, where P and Q are abstract terms which are usually represented by letters and numbers. The idea is to see whether participants can derive the syllogism by evaluating both modus ponens ($P \therefore Q$) and modus tollens ($\neg Q \therefore \neg P$) of the premise.

The most common form of this task is to present participants with a set of cards which always has a letter on one side and a number on the other (illustrated in Figure 5.1). Four cards are laid out in front of the participant depicting a vowel (P), a consonant ($\neg P$), an even number (Q), and an odd number ($\neg Q$). The participant is then provided with the premise: ‘if a card has a vowel on one side, then it has an even number on the other side’. The task is to tell the experimenter which card or cards *must* be turned over in order to expose the validity of the premise.

The correct answer is to turn the card with the vowel (P ; evaluates modus ponens) and the card with the odd number ($\neg Q$; evaluates modus tollens). If there is not an even number on the other side of the vowel card or if there is a vowel on the other side of the odd number card, the premise has been falsified. Conversely, if there is an even number on the other side of the vowel card *and* if there is not a vowel on the odd number card, the premise has been verified. For the truth and falsity of the premise, the print on the other side of the two remaining cards make no difference whatsoever from a formal logic point of view.

What participants tend to do in this task, however, is to either suggest only the vowel (P) card or the vowel (P) card and the even number (Q) card. But, the premise does not say anything about what a consonant card might have on the other side which means that turning the even number card and finding a consonant does not falsify the premise.

Providing Context

In 1971, Wason and Shapiro replicated the the original study showing no results of participants logical reasoning even when they were given practice rounds of logical syllogisms beforehand. However, in the study they did a variation where participants were given the rule: ‘every time I go to Manchester I travel by car’ and the four cards now consisted of a form of transportation on one side and a city name on the other. In this version, participants were more likely to provide the modus tollens ($\neg Q$) response.

Griggs and Cox (1982) devised a version of Wason’s selection task which provides a theme familiar to the participants. In this version, the cards have an age on one side and a type of drink on the other. The participant is then provided with this premise instead: ‘if a person is drinking beer, then the

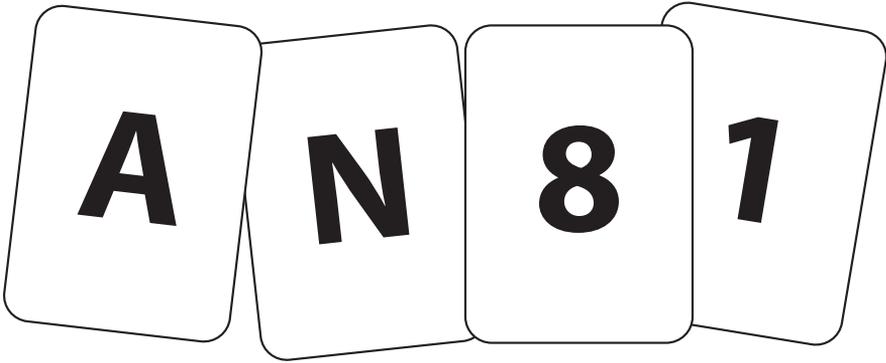


Figure 5.1: An example of the setup in Wason's selection task

person must be over 19 years of age'. Participants are also instructed that they are police officers whose job is to make sure people abide to this rule. In this version of the task, the majority of participants were able to solve the syllogism correctly.

Evans (1989) made a thorough review of different versions of the task available in the literature to that day. He makes a distinction between form, content, and context. Form is the abstracted, formal logic of the task whilst content is the parts we fill the task with (e.g., letters and numbers or ages and drinks). Context he consider to be the background information or experimental instructions related to the content. If the context, in his terms, is something the participant is familiar with, the task becomes easier to accomplish.

There are several examples where experiments of reasoning, decision-making, and problem-solving have been reproduced with altered instructions resulting in mitigation of cognitive biases. For example, the conjunction fallacy can be mitigated by asking people to reason by frequencies rather than probability. In the Linda problem, if asked to answer how many out of 100 people matching Linda's description would match with each of the provided options, participants tend to provide a larger frequency for the option of being only an active feminist than being both a bank teller and an active feminist (Fiedler, 1988).

What this shows is the importance of relevant information available for a problem-solver to formulate a context in which the problem can be solved more effectively. What this implies for the possibility of a designer to construct a workable system is that relevant information must be obtained from a variety of target environments in order to form a valid representation of the problem domain. How problems are represented is crucial for the possibility

of solving problems because it has a strong influence on the problem-solvers cognition (Zhang & Norman, 1994).

For designers, this implies that they need to be able to formulate a representative context of a problem domain by getting close to target environments to elicit valuable information in order to build a useful and usable system. Sometimes, however, target environments might not be within reach. The designer then has to find methods to get as close to the problem domain as possible.

Developing for a Future World (Paper III)

When developing systems in completely new domains or domains which are being revolutionised, target environments are unknown territory. In working with developing a human-machine interface for an automated long-haul truck (Krupenia et al., 2014), me and my colleagues ran into issues when trying to understand possible strategies that a future driver might use to manage everyday working tasks. In order to find possible work strategies, one has to understand the work domain. However, when it comes to future systems which revolutionises the work domain, the impact the new system has on the work domain is difficult to anticipate. In these circumstances, researchers are at risk of ending up in a ‘catch 22’ situation where the work domain is needed in order to develop a product but in order to understand the impact the product has on the work domain, the product itself must exist in some form.

Strategies Analysis

Throughout the project we employed the Cognitive Work Analysis (CWA) framework (Vicente, 1999). We modelled the current work domain of cargo delivery with long-haul trucks throughout the initial stages of the framework. However, when we came to what is called the Strategies Analysis phase, we suddenly hit a road bump when trying to weed out strategies for performing tasks identified in the previous work with the framework. The idea of the strategies analysis is to elicit as many safe strategies as possible for performing a certain task. This in order to reduce possibilities of the future system to constrain the operators as a means to facilitate system resilience in case of detrimental contingencies.

The problem we encountered was that the future system is meant to be a highly automated, self-driving vehicle. This means that, even though there is little difference in the overall tasks that the entire socio-technical system is meant to perform in the future compared to today, responsibility to fulfil these tasks is shifting drastically when high automation is introduced. Thus,

the operators' work will consist more of monitoring and less of driving. To further complicate a strategies analysis, handover and takeover of driving and other tasks should be flexible so that the operator can assume control in case of an emergency.

Formative Approach

To follow through with the analysis, we decided to adopt a formative strategies analysis devised by Hassall and Sanderson (2014). This method is a means to envision work strategies by using domain-independent categories based mainly on knowledge of human cognition. The authors provide strategy categories identified through an extensive literature review of works related to human cognition. The authors' categories were (1) intuitive, (2) analytical reasoning, (3) compliance, (4) option-based, (5) imitation, (6) avoidance, (7) cue-based, and (8) arbitrary-choice. These categories are then meant to be populated with as many strategies as possible for tasks discovered through previous activities in the CWA framework. Once these strategies are defined, the researcher tries to identify factors that may affect these strategies as well as factors that prompts each strategy.

In our project, we identified strategies through three workshops with a total of 15 long-haul truck drivers using the 8 strategy categories provided by Hassall and Sanderson. The workshops were conducted in front of a whiteboard where we projected familiar scenarios from truck driving, drawing interface ideas on these projected scenarios, contemplating modifications to the interface whilst discussing strategies of various tasks to be performed with a highly automated truck. We discussed present strategies to achieve work tasks and jointly elaborated on possible future strategies.

Conclusion

If you consider the experimenter trying to condition a rodent in a laboratory as discussed in Chapter 3, it is the experimenter that formulates the context. The experimenter considers light and lever to be the only stimulus relevant in the experiment. To the rat, however, everything available in its environment is a stimulus. This is a crucial illustration of the difference between context and environment. As a designer when you are in your design 'lab', you are not much different from the experimenter.

When information is scarce, the formulation of context is largely in the hands of your personal biases. The 2-4-6 task shows the importance of posing the right questions and the need to falsify ones hypotheses (Popper, 1959) in order to overcome confirmation bias and reveal crucial information. This is highly relevant for any designer. Woods (1998) has argued for the need

to treat any design as a hypothesis since “products and prototypes embody hypotheses about what would be useful” (p. 168). Woods points to the representation effect on problem-solvers cognition as exemplified through empirical findings discussed in the present chapter and how this then influences the ability to use a particular product or system depending on its design.

In our work within the MODAS project discussed above, we had trouble getting to an actual user environment. This was handled by using actual drivers for knowledge elicitation as well as constructing a simulator in which we could get as close to a future world as possible. In our strategies analysis we also used images representing difficult driving environments such as steep hills or highly trafficked motorways. These stimuli are of course lacking in many modalities which would normally be present in reality, but the simulator and scenario images should at least help drivers think about more concrete situations, making them more able to reason about what would be considered helpful design solutions for a future, highly automated vehicle.

In the title of Paper III, we refer to ‘context-free strategy categories’. Now, the question here is whether anything can actually be context-free. What we were having trouble with in the project was the ability to formulate representative contexts due to the fact that we could not observe or interview any user to get to any actual target environments. I would rather consider that the strategy categories we employed were domain independent since I do not think that a problem-solver can actually think of a problem without a context. If there is not enough environmental information to base a context on, information stored in memory will be utilised instead aided by biases in formulating a workable context leading the problem-solver to strategies for producing a solution.

Chapter 6

Discussion

“Something is context because of the way it is used in interpretation, not due to its inherent properties”

Winograd (2001, p. 405)

A theory of human-computer interaction is nothing short of a theory of interaction with reality. Understanding interaction will be a crucial key in understanding phenomenal consciousness. With noumenal reality we negotiate borders of interaction, I contend that there is no difference between human-computer interaction and human-human interaction or even human-cloth interaction. We make a distinction in the interaction only because with some noumenal objects, the negotiation of rules how to interact is ever on-going, whilst with other noumenal objects the negotiation is finished before it even starts. When interacting with a piece of cloth, I early establish the boundaries of my interaction because this noumenal object’s behaviour is easily theorised and thus it behaves consistently with expected hypotheses of its use. Humans, on the other hand, are not as consistent in how they behave towards my actions (be they words or punches) and thus we constantly renegotiate our interactive boundaries. Now, computers is a special sort of interaction only because it is in the realm between interacting with a cloth and a human. A computer is in some respects as advanced as humans but as communicative as cloths which makes interpretation of current states impenetrable and that has detrimental effects on this communicative negotiation.

Finding and defining precise terms is a necessity for a proper nomenclature within a scientific field and the foundation of clear communication. I had a seminar on context recently where one of the participants of the seminar said that we have a better word than ‘context’ in Swedish: what

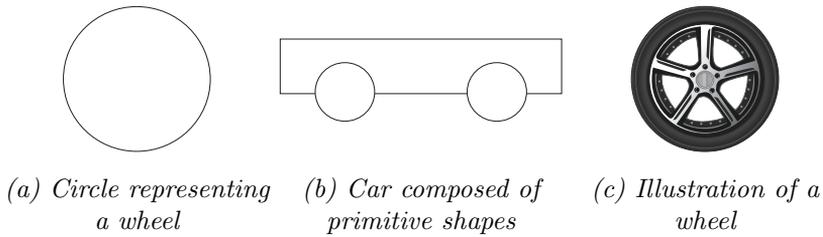


Figure 6.1: Different abstraction levels of a wheel

we call ‘sammanhang’ (from the German word ‘Zusammenhang’). This word gives Swedish speakers a more intuitive sense of what the word is meant to describe, because its constituents afford the use of the term. ‘Sammanhang’ is a compound of ‘together’ and ‘hanging’ (i.e. concepts are ‘hung together’), you instinctively know what the speaker means by hearing the term because you understand its constituents. The term ‘context’, on the other hand, does not afford the same intuitive sense to a Swede because it does not offer the same intuitive sense of its constituents since the Swedish language root is Germanic and not Italic.¹ However, to someone familiar with the origin of the term ‘context’ (i.e., derived from manufacturing textiles by combining individual threads to form a whole larger than the sum of its parts) you realise that the terms have exactly the same connotations and this will make understanding of the term clearer. In this sense, learning about its Latin origin helps you develop the context of the term itself. Context is thus crucial for formation of memories, because it tells you in what box to categorise a newly learnt fact, whilst incoming, raw and noisy data is what chooses the boxes stored in your memory and make them available to you.

Who Provides the Context?

This is why it is so easy for us to substitute the map for the territory and think that context is part of the noumenal reality and thus something situated in our environments. Take a look at Figure 6.1a. What do you see? Well nothing more than a circle you might argue. Take a look at only the circle in Figure 6.1b, what do you see? All people I have asked thus far consider this circle to be a wheel. In trying pointing to a specific part in this picture I invoke a problem-solving goal for you. Now what you probably

¹The word ‘germane’ is an interesting example here. It means ‘relevant to a subject under consideration’ where the root is obviously the same as the word for the country Germany. I suspect that this has to do with the fact that English is by root a Germanic language. In this tautological word lies the meaning that it is in the German language (context) you have to look in order to find meaning to the use of some words.

saw first was a car. I have tried to show this picture to colleagues and a few times people instinctively say: 'it's a car', even though I specifically and distinctly only point to the circle in the image by tracing it with my finger whilst asking what it is. You see, it is impossible not to consider all incoming data as a whole and once you have the whole you can determine the parts. When the whole is learnt it cannot be un-learned. This means that a circle found in that particular position on what resembles a car (or a wagon, or a trailer) must reasonably be a wheel. We have long since established that context matters and this example (and the examples of ambiguous symbols in Figure 2.1) makes it easy to conclude that context is something we take in from the environment. But consider then Figure 6.1c. Is it not obvious that this is a wheel? If we use a standard definition of 'context' that you place something in focus and everything around it is context. How does context help you here? There are no cars, no traffic, very little which we relate to a 'wheel context' on this page. There is nothing but white paper and printed ink surrounding the wheel, so how do you know it is a wheel?

Well, it is not that which is around it in noumenal reality which is important, but that which is around it in your phenomenal reality. Context is all ideas and concepts mentally connected to the incoming stimuli. So why do you not instinctively think of a circle when admiring Figure 6.1c? That is because context is not that which reality offers you, it is that which you bring to the table. When you see the wheel, you will be bringing in memories that fit the incoming data in relation to previous experiences. You have only ever seen this phenomenal object (or symbol representing the specific noumenal object) on vehicles, most often cars. You will automatically think of a car wheel, because the symbol is strongly coupled with the concept of cars. A circle, on the other hand, you have seen in relation to a multitude of noumenal objects, and therefore you can be very creative in choosing how to interpret this symbol in deciding which context to relate it to. Thinking of the metaphor of context as a box: a specific wheel fits in one box which in turn is placed in the box of circles containing many boxes of circular, phenomenal objects.

Goal-Dependence

Alfred Yarbus was pioneering in the field of eye tracking. In his now seminal work on top-down, cognitive processing he presented participants with a picture of the painting *An Unexpected Visitor* by Ilya Repin. He then gave participants verbalisation tasks such as 'give the ages of the people shown in the picture' and 'surmise what the family was doing before the arrival of the unexpected visitor' (Yarbus, 1967). Each task resulted in distinct

eye movement patterns in the participants. When asked to give the age of the portrayed individuals, fixations were essentially on nothing but the individuals' faces whilst when explicating what had been happening before the arrival of the visitor, participants tended to fixate on objects drawn on the table in the painting. Yarbus voices that "during perception many of the elements of the picture are not perceived by foveal vision. [...] Foveal vision is reserved mainly for those elements containing essential information needed by the observer during perception." (Yarbus, 1967, p. 196).

In collecting information about the individuals' ages, very few re-fixations on faces were observed. However, when given the task 'estimate how long the unexpected visitor has been away from the family', saccades were reiterated and fixations made many times on faces. What I suspect is happening is that participants are homing in on the answer in interpreting facial expressions through a sort of hypotheses formulation and testing procedure. By considering different time lengths as underlying hypotheses, the observer can perform hypothesis testing against supposed relevant facial data in order to assess the validity of each hypothesis of how long the visitor has been away.

What this study of process-tracing illustrates is that as humans we are goal-oriented in how we collect information from available cues in our environments. We do not haphazardly take in any and all data, but instead focus our attention on what we consider relevant, perceptual chunks of the environment. In this study, I would argue that people formulated contexts in which the tasks could be solved by purposely focusing attention on what they considered relevant for each posed problem. These findings points to the necessity of a goal for context formulation.

Foundation of Attention

Context, in colloquial sense, really only means 'entirety'. Context is a model of reality, if a context is perfectly implemented it is no longer a model but reality itself. Context can thus be defined as a mental model of relevant aspects of an environment given a specified goal. When acting directly with the environment, we do so from pre-conceived expectations on the environment. When these expectations are not met, we need to recreate events to understand where the discrepancy occurred. This recreation of environmental events is a dynamic, mental process; a simulation which builds up what we call 'context'. Thus, we are not part of a context but instead we are mentally situated in the environment, but we have the ability to mentally simulate events using concepts we have encountered in the environment and stored in long-term memory. Furthermore, when we are engaging with re-

ality, context helps us extract vital information present. This has been well illustrated in experiments of inattention blindness (Mack & Rock, 1998). In the most famous experiment (Simons & Chabris, 1999), participants are watching a video of two teams of people in either black or white t-shirts and asked to keep track of aerial and bounce passes made within one of the teams. In the video, a woman appears either with an umbrella or dressed in a gorilla suit. When mentally engaged in the counting task 50 % of the participants did not take notice of the woman. This suggests that participants have formulated a specific task context in which vital data from the surrounding is made perceptually available in relation to the stipulated task goal and the rest is pretty much ignored.

My understanding is that context is not something we are a part of, but rather something we create mentally in relation a specific goal. Determination of something ambiguous thus comes from top-down processes related to a goal. I believe context has been wrongly interpreted in HCI as that which a user is situated *in* and which a product is being used *in*. I suggest instead a separation between the user environment and the user context. The user environment is what a specific individual is situated in and that which we would like to understand. A context on the other hand is something which human problem-solvers instantiate, in the case of designers it would be a generalised, mental representation of the user environments (i.e., a usage domain) which guides designers to make informed design decisions. Contextualisation is thus a generalised function of all intelligent agents able to solve problems; I view it as a fundamental part of problem-solving and decision making.

Context, in this sense, is fundamental to attention. Context facilitates making sense of unstructured data, building comprehensible percepts out of ambiguous sense data. Context can be viewed as a fabric of concepts determining the boundaries of what concepts to be made available mentally. These intertwined concepts are imprints of real, experienced noumena forming a base for what to expect in engagement with reality. Context can thus be viewed similarly to a script for your behaviour; in this sense context is a loosely defined algorithm driving attention and action. When a goal is formulated then relevant concepts have been woven together and through this conceptual fabric, both mental simulations and behaviour can be driven in a sort of suspended algorithm which is dynamically executed and re-written on the fly in response to environmental fluctuations. Context therefore facilitates interaction between mind and matter in bridging top-down and bottom-up processes. Problem formulation is the way in which a context is mentally loaded and the task goal is that which determines what details will be made consciously available through perceptual processing of chaotic sensory data.

nicant

Figure 6.2: An incomprehensible word

Reifying Context for the Reader

The ideas of Baars (1988) with regard to context, is the closest to the way I interpret the term with a slight modification. I do not agree with that context is something we are not conscious of. Granted, the processes by which reality presents the context is unknown to me, but the resulting effect is very much consciously available. We make these contexts available by collecting data available in reality. A context is just as important to a user as it is to a developer. In order to fully comprehend possible user tasks, we must immerse ourselves in the domain of our user which allows for context formation.

Context can be considered to be the colloquial term for the sensation of having a neural network and the folk psychological theory of memory formation we could call contextualisation (i.e., data interpretation). Context is a memory structure; when you have no context, you cannot formulate hypotheses of how incoming, ambiguous data should be categorised. But, as soon as you have a context, interpretation becomes easier because it fits with previously collected data and you can fit this new data in the structure which let you formulate your hypothesis. Let me illustrate with an example. Take your time and study what is illustrated in Figure 6.2. What is it? The closest thing you can think of that it remotely resembles is a word. But which word is it? In using a metaphor, you open different boxes to see where this incoming data fits. You found the box of words and it seems to fit there, but is it really a word and in that case what words fit. Does a word such as ‘nicant’, for example, exist? There sure is something called ‘picant’, but is that first letter really a ‘p’?

Now, take a look at Figure 6.3. What does it say? It is not very easy to decipher but I am still to come across a person who is not able to read and comprehend this excerpt from a sentence. Now look back at Figure 6.2, is the incoming data easier to interpret now? Once you have learnt to interpret data, this memory cannot be undone. You spent time looking at the word and forming hypotheses, these hypotheses fly out the window as you experienced this data together with other incoming data that together makes sense; it fits. This will help you interpret what is printed in Figure 6.2 for a good part of you life.

The more experience you gain with similar noumena the quicker you can categorise incoming phenomenal data and the more automated your behaviour becomes. When reading the excerpted phrase in Figure 6.3, the

word just popped out because you made it available through previous knowledge of how the word is usually paired with other words. When you then went back and looked at only the word in Figure 6.2, these newly formed structures instantly led you to interpret the word. Activation of more relevant neural structures happens as more data accumulates and finally you strike proverbial gold. This example is as pure a phenomenal experience of memory formation and recall if ever there was one.

Conclusion

Do we really need HCI? Do people not learn systems just by being exposed to them? Context could be likened with ‘categorised experiences’. Building context can be considered the acquisition of richer concepts, describing experience with words, which is the same as what we call knowledge or as I prefer it, discovery of the noumenal world. I think with HCI we have created a non-problem. People learn how to use technology if they value the goals it delivers. Learning is just adapting to noumenal objects and therefore usability might not really be the real problem, but instead it might be peoples fears of not succeeding in the interaction.

were meant to describe

Figure 6.3: Comprehensible phrase

Epilogue

My experiences are different from yours, but by communicating how I experience something, we together formulate an pseudo-objective view of reality by collecting data from each others perceptual experiences of noumenal reality. I have, in this thesis, exposed my ignorance for you to scrutinise. I am fully aware that I have taken the concept of context to an extreme and the only way for me to learn is to see how this idea is handled in the scientific community in order to expose its flaws. By doing that, a negotiation starts between me and my peers where other extreme views of the concept helps me create the borders of the concept in a conceptual space. By getting feedback in working towards the goal of pinning down a concept, I adjust my view and you adjust yours in a discriminatory process which lets me know what notions to stay with and what notions to refute. As a metaphor, you can consider ideas to be cells which divide themselves to form a structure, when talking about ideas and how to categorise them, a conceptual structure formulates. This is why 'stupid questions' are fundamental to learning because it propels learning faster in that you immediately trim your notion of a an idea vastly when being educated by someone's response. The more I understand the world, the more I can seize opportunities in reality.

To perceive is to gain knowledge, and action is to practice theories in order to understand. This means that to teach is to gain understanding whilst to learn is acquire knowledge. Knowledge has to do with trimming motor behaviour to incoming perceptual experiences in an automation process. Understanding has to do with prediction of reality, it is our cognitive, top-down processes testing hypotheses drawn from our knowledge source of perceptual memories gained through knowledge. The ultimate goal of living and the meaning of life, for me, is learning. Learning is the method of life where survival is the goal. Learning I define as adaptation to reality, and the goal is to resonate with reality because that is how I define survival. An organism incapable of resonating in crucial aspects with reality is destined for a dirt nap. If you are afraid to be stupid, then you are also afraid to learn. When we do not dare to expose our ignorance we become victims to ignorance which is the suffocation of knowledge and death of understand-

ing. For long, we have got the relationship between teacher and student only partially right because it is not a one-way relationship. As students we gain knowledge from teachers and as teachers we gain understanding from students; all in the institution of answering 'stupid questions' that we call school. It is in the realm where theory interacts with practice that learning takes place and in which consciousness thrives. A quiet classroom is a waste land; a non-fertile ground in which seeds of knowledge cannot grow.

Acknowledgments

I would like to thank my collaborators at Lund University for a very interesting and rewarding project of educational technology in preschool. Thanks go to Agneta Gulz and Richard Andersson who supervised the project and were tremendously helpful in developing my experimental and statistical thinking. Thanks also go to Mette Clausen-Bruun and Magnus Haake for being valuable assets in this project.

I would like to thank my collaborators at Scania for involving me in the MODAS project. Thanks go to Stas Krupenia, Johanna Vännström, Victor Ahlm, and Stefan Larsson.

I aim a special thank you to my colleague and friend Mikael Laaksoharju for being there and listening to my ideas and, together also with thanks you Iordanis Kavathatzopoulos, for helping me develop my critical thinking.

I also wish to thank my PhD student colleagues who put up with my arrogance: Ida Bodin, Ted White, Thomas Lind, and also Simon Tschirner (when are you coming back to empty the shelf so I can put my books up?). And also, a great many thank you to those of you who were able to participate in my seminar on context during the spring.

This thesis has been through numerous revisions, constantly improving. I have my three brilliant supervisors to thank for this enlightening process: Anders Jansson, my main supervisor and good friend who constantly offers a shoulder to cry on and listens to my conceited ideas; Mats Lind, my co-supervisor and sounding board who always throws my ideas back at me and understands what I mean, not what I am saying; and finally, Annika Wallin, my second co-supervisor and sun who constantly melts my wings and brings me back to earth. Thanks to the three of you for making this process a delight!

References

- Akman, V. (2002). Contesti in intelligenza artificiale: una fugace rassegna. In C. Penco (Ed.), *La svolta contestuale milano* (pp. 147–166). McGraw-Hill. (English: Context in artificial intelligence: A fleeting overview)
- Anderberg, E., Axelsson, A., Bengtsson, S., Håkansson, M., & Lindberg, L. (2013). Exploring the use of a teachable agent in a mathematical computer game for preschoolers. *Lund University Cognitive Studies*, *154*, 161–171.
- Axelsson, A., Anderberg, E., & Haake, M. (2013). Can preschoolers profit from a teachable agent based play-and-learn game in mathematics? *Lecture Notes in Computer Science*, *7926*, 289–298.
- Baars, B. J. (1988). *A cognitive theory of consciousness*. Cambridge, UK: Cambridge University Press.
- Bargh, J. A. & Schul, Y. (1980). On the cognitive benefits of teaching. *Journal of Educational Psychology*, *72*, 583–604.
- Bevan, N. & Macleod, M. (1994). Usability measurement in context. *Behaviour and Information Technology*, *13*, 132–145.
- Beyer, H. & Holtzblatt, K. (1998). *Contextual design: defining customer-centered systems*. London, UK: Academic Press.
- Biswas, G., Leelawong, K., Schwartz, D. L., Vye, N., & The Teachable Agents Group at Vanderbilt. (2005). Learning by teaching: A new agent paradigm for educational software. *Applied Artificial Intelligence*, *19*, 363–392.
- Blair, K., Schwartz, D. L., Biswas, G., & Leelawong, K. (2006). Pedagogical agents for learning by teaching: Teachable agents. *Educational Technology & Society*. (Special Issue on Pedagogical Agents.)
- Brézillon, P. (1996). *Context in human-machine problem solving: A survey* (Tech. Rep.). Paris, France: LAFORIA.
- Cacioppo, J. T., Priester, J. R., & Berntson, G. G. (1993). Rudimentary determinants of attitudes. ii: Arm flexion and extension have differential effects on attitudes. *Journal of Personality and Social Psychology*, *65*, 5–17.

- Cawdrey, R. (1604). *A table alphabetically containing and teaching the true writing, and understanding of hard vsuall english wordes*. London, UK: Mathew Simons.
- Chaiklin, S. & Lave, J. (Eds.). (1993). *Understanding practice: Perspectives on activity and context*. Cambridge, UK: Cambridge University Press.
- Chalmers, M. (2004). A historical view of context. *Computer Supported Cooperative Work (CSCW)*, 13, 223–247.
- Chase, C. C., Chin, D. B., Opezzo, M. A., & Schwartz, D. L. (2009). Teachable agents and the protégé effect: Increasing the effort towards learning. *Journal of Science Education and Technology*, 18, 334–352.
- Chen, H., Finin, T., & Joshi, A. (2003). An ontology for context-aware pervasive computing environments. *The Knowledge Engineering Review*, 18, 197–207.
- Clark, A. & Chalmers, D. (1998). The extended mind. *Analysis*, 58, 7–19.
- Clark, H. H. (1996). *Using language*. Cambridge, UK: Cambridge University Press.
- Clark, H. H. & Carlson, T. B. (1981). Context for comprehension. In J. Long & A. Baddeley (Eds.), *Attention and performance IX* (pp. 313–330). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Clements, W. A. & Perner, J. (1994). Implicit understanding of belief. *Cognitive Development*, 9, 377–395.
- Crowley, J. L., Coutaz, J., Rey, G., & Reignier, P. (2002). Perceptual components for context aware computing. In *Ubicomp 2002: Ubiquitous computing* (pp. 117–134). Springer.
- Dey, A. K. (2001). Understanding and using context. *Personal and Ubiquitous Computing*, 5, 4–7.
- Dey, A. K. & Abowd, G. D. (2000). Towards a better understanding of context and context-awareness. In *Proceedings of the what, who, where, when, and how of context-awareness workshop of chi*. New York, NY: ACM Conference on Human Factors in Computer Systems.
- Dourish, P. (2004). *Where the action is: the foundations of embodied interaction*. Cambridge, MA: MIT press.
- Evans, J. S. B. T. (1989). *Bias in human reasoning: Causes and consequences*. East Sussex, UK: Lawrence Erlbaum Associates.
- Fiedler, K. (1988). The dependence of the conjunction fallacy on subtle linguistic factors. *Psychological Research*, 50, 123–129.
- Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist*, 34, 906–911.
- Frankland, P. W., Cestari, V., Filipkowski, R. K., McDonald, R. J., & Silva, A. J. (1998). The dorsal hippocampus is essential for context discrimination but not for contextual conditioning. *Behavioral Neuroscience*,

- 112, 863–874.
- Fukushima, J., Hatta, T., & Fukushima, K. (2000). Development of voluntary control of saccadic eye movements i. age-related changes in normal children. *Brain & Development*, 22, 173–180.
- Gay, J. & Cole, M. (1967). *The new mathematics and an old culture : a study of learning among the kpelle of liberia*. New York, NY: Holt, Rinehart and Winston.
- Godden, D. R. & Baddeley, A. D. (1975). Context-dependent memory in two natural environments: On land and underwater. *British Journal of Psychology*, 66, 325–331.
- Goodwin, C. & Duranti, A. (1992). Rethinking context: an introduction. In C. Goodwin & A. Duranti (Eds.), *Rethinking context: Language as an interactive phenomenon* (pp. 1–42). Cambridge, UK: Cambridge University Press.
- Grainger, J. & Whitney, C. (2004). Does the human mind read words as a whole? *Trends in Cognitive Sciences*, 8, 58–59.
- Griggs, R. A. & Cox, J. (1982). The elusive thematic materials effect in wason's selection task. *British Journal of Psychology*, 73, 407–420.
- Gulz, A., Haake, M., & Silvervarg, A. (2011). Extending a teachable agent with a social conversation module – effects on student experiences and learning. *Artificial Intelligence in Education*, 6738, 106–114.
- Haake, M., Axelsson, A., Clausen-Bruun, M., & Gulz, A. (2015). Scaffolding mentalizing via a play-&-learn game for preschoolers. *Computers & Education*, submitted.
- Hall, G. & Honey, R. C. (1989). Contextual effects in conditioning, latent inhibition, and habituation: Associative and retrieval functions of contextual cues. *Journal of Experimental Psychology: Animal Behavior Processes*, 15, 232–241.
- Hallett, P. E. (1978). Primary and secondary saccades to goals defined by instructions. *Vision Research*, 18, 1279–1296.
- Hassall, M. E. & Sanderson, P. M. (2014). A formative approach to the strategies analysis phase of cognitive work analysis. *Theoretical Issues in Ergonomics Science*, 15, 215–261.
- Henricksen, K., Indulska, J., & Rakotonirainy, A. (2002). Modeling context information in pervasive computing systems. In F. Mattern & M. Naghshineh (Eds.), *Pervasive computing* (pp. 167–180). Berlin Heidelberg: Springer-Verlag.
- Hohwy, J. (2013). *The predictive mind*. Oxford, UK: Oxford University Press.
- Hollan, J., Hutchins, E., & Kirsh, D. (2000). Distributed cognition: Toward a new foundation for human-computer interaction research. *ACM Transactions on Computer-Human Interaction*, 7, 174–196.

- Hughes, C. (1998). Executive function in preschoolers: Links with theory of mind and verbal ability. *British Journal of Developmental Psychology*, *16*, 233–253.
- Hutchins, E. (1995). How a cockpit remembers its speeds. *Cognitive Science*, *19*, 265–288.
- Hutchins, E. & Klausen, T. (2000). Distributed cognition in an airline cockpit. In Y. Engström & D. Middleton (Eds.), *Cognition and communication at work* (pp. 15–34). New York, NY: Cambridge University Press.
- ISO. (1998). *Ergonomic requirements for office work with visual display terminals (VDTs) — part 11: Guidance on usability* (ISO No. 9241-11). Geneva, Switzerland: International Organization for Standardization.
- Kant, I. (1922/1781). *Critique of pure reason*. London, UK: The MacMillan Company.
- Kirsh, D. & Maglio, P. (1994). On distinguishing epistemic from pragmatic action. *Cognitive Science*, *18*, 513–549.
- Klein, G. A. (1993). A recognition-primed decision (rpd) model of rapid decision making. In G. A. Klein, J. Orasanu, R. Calderwood, & C. E. Zsombok (Eds.), *Decision making in action: Models and methods* (pp. 138–147). Norwood, NJ: Ablex Publishing Corporation.
- Klein, G. A., Calderwood, R., & Clinton-Cirocco, A. (2010). Rapid decision making on the fire ground: The original study plus a postscript. *Journal of Cognitive Engineering and Decision Making*, *4*, 186–209.
- Krupenia, S., Selmarker, A., Fagerlönn, J., Delsing, K., Jansson, A., Sandblad, B., & Grane, C. (2014). The ‘Methods for Designing Future Autonomous Systems’ (MODAS) project. In N. Stanton, S. Landry, G. D. Bucchianico, & A. Vallicelli (Eds.), *Advances in human aspects of transportation: Part II* (pp. 70–81). AHFE Conference.
- Kull, K. (2010). Umwelt. In P. Cobley (Ed.), *The routledge companion to semiotics* (pp. 348–349). London, UK: Routledge.
- Lakoff, G. & Johnson, M. (1980). *Metaphors we live by*. Chicago, IL: The University of Chicago Press.
- Lave, J. (1988). *Cognition in practice: mind, mathematics and culture in everyday life*. Cambridge, UK: Cambridge University Press.
- Lee, J.-E. R., Nass, C. I., & Bailenson, J. N. (2014). Does the mask govern the mind?: Effects of arbitrary gender representation on quantitative task performance in avatar-represented virtual groups. *Cyberpsychology, Behavior, and Social Networking*, *17*, 248–254.
- Loke, S. W. (2004). Representing and reasoning with situations for context-aware pervasive computing: a logic programming perspective. *The Knowledge Engineering Review*, *19*, 213–233.
- Mack, A. & Rock, I. (1998). *Inattentional blindness*. Cambridge, MA: MIT

- press.
- Maguire, E. A., Woollett, K., & Spiers, H. J. (2006). London taxi drivers and bus drivers: A structural mri and neuropsychological analysis. *Hippocampus*, *16*, 1091–1101.
- Maister, L., Slater, M., Sanchez-Vives, M. V., & Tsakiris, M. (2015). Changing bodies changes minds: owning another body affects social cognition. *Trends in Cognitive Sciences*, *19*, 1091–1101.
- McDermott, R. P. (1993). The aquisition of a child by a learning disability. In S. Chaiklin & J. Lave (Eds.), *Understanding practice: Perspectives on activity and context* (pp. 269–305). Cambridge, UK: Cambridge University Press.
- McGroch, J. A. (1939). Learning. In E. G. Boring, H. S. Langfeld, & H. P. Weld (Eds.), *Introduction to psychology*. New York, NY: Wiley.
- Munoz, D. P. & Everling, S. (2004). Look away: The anti-saccade task and the voluntary control of eye movement. *Nature*, *5*, 218–228.
- Nardi, B. A. (1995). Studying context: a comparison of activity theory, situated action models, and distributed cognition. In B. A. Nardi (Ed.), *Context and consciousness: activity theory and human-computer interaction* (pp. 69–817). Cambridge, MA: MIT Press.
- Ogan, A., Finkelstein, S., Mayfield, E., D’Adamo, S., Matsuda, S., & Caspell, J. (2012). “oh,dear stacy!” social interaction, elaboration, and learning with teachable agents. In *Proceedings of the sigchi conference on human factors in computing systems* (pp. 39–48). ACM.
- Okoli, J. O., Weller, G., & Watt, J. (2016). Information processing and intuitive decision-making on the fireground: towards a model of expert intuition. *Cognition, Technology & Work*, *18*, 89–103.
- Pareto, L., Haake, M., Lindström, P., Sjödén, B., & Gulz, A. (2012). A teachable-agent-based game affording collaboration and competition: Evaluating math comprehension and motivation. *Educational Technology Research and Development*, *60*, 723–751.
- Pascoe, J. (1998). Adding generic contextual capabilities to wearable computers. In *Wearable computers, 1998. digest of papers. second international symposium on* (pp. 92–99).
- Peirce, C. S. (1905). What pragmatism is. *The Monist*, *15*, 161–181.
- Perner, J. & Lang, B. (1999). Development of theory of mind and executive control. *Trends in cognitive sciences*, *3*, 337–344.
- Popper, K. (1959). *The logic of scientific discovery*. New York, NY: Basic Books.
- Rayner, K., White, S. J., Johnson, R. L., & Liversedge, S. P. (2006). Raeding wrods with jubmled lettres. *Psychological Science*, *17*, 192–193.
- Reiff, R. & Scheerer, M. (1959). *Memory and hypnotic age regression: Developmental aspects of cognitive function explored through hypnosis*.

- New York, NY: International Universities Press.
- Reignier, P., Brdiczka, O., Vaufreydaz, D., Crowley, J. L., & Maisonnasse, J. (2007). Context-aware environments: from specification to implementation. *Expert systems*, *24*, 305–320.
- Robins, P. & Aydede, M. (Eds.). (2008). *The cambridge handbook of situated cognition*. New York, NY: Cambridge University Press.
- Rodden, T., Cheverst, K., Davies, K., & Dix, A. (1998). Exploiting context in hci design for mobile systems. In *Workshop on human computer interaction with mobile devices* (pp. 21–22).
- Schilit, B. N., Adams, N., & Want, R. (1994). Context-aware computing applications. In *Mobile computing systems and applications, 1994. wmcasa 1994. first workshop on* (pp. 85–90).
- Schilit, B. N. & Theimer, M. M. (1994). Disseminating active map information to mobile hosts. *Network, IEEE*, *8*, 22–32.
- Schmidt, A., Beigl, M., & Gellersen, H.-W. (1999). There is more to context than location. *Computers & Graphics*, *23*, 893–901.
- Schwartz, D. L., Chase, C. C., Chin, D. B., Opezzo, M. A., Kwong, H., Okita, S., . . . Biswas, G. (2009). Interactive metacognition: Monitoring and regulating a teachable agent. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Handbook of metacognition in education* (pp. 340–358). Routledge Press.
- Scribner, S. & Cole, M. (1973). Cognitive consequences of formal and informal education. *Science*, *182*, 553–559.
- Shephard, R. N. & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, *171*, 701–703.
- Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review*, *63*, 129–138.
- Simon, H. A. (1996). *Sciences of the artificial* (3rd ed.). Cambridge, MA: MIT press.
- Simons, D. J. & Chabris, C. F. (1999). Gorillas in our midst: sustained inattentive blindness for dynamic events. *Perception*, *28*, 1059–1074.
- Smith, D. C., Irby, C., Kimball, R., & Harslem, E. (1982). The star user interface: An overview. In *Proceedings of the national computer conference* (pp. 515–528). New York, NY, USA: ACM.
- Smith, S. M., Glenberg, A., & Bjork, R. A. (1978). Environmental context and human memory. *Memory & Cognition*, *6*, 342–353.
- Thomas, C. & Bevan, N. (Eds.). (1996). *Usability context analysis: a practical guide*. Teddington, UK: NPL Usability Services.
- Turner, R. M. (1993). Context-sensitive reasoning for autonomous agents and cooperative distributed problem solving. In *Proceedings of the ijcai workshop on using knowledge in its context* (pp. 141–151).

- Tversky, A. & Kahneman, D. (1981). *Judgments of and by representativeness* (Tech. Rep.). Arlington, VA: DTIC Document.
- Varela, F. J., Thompson, E., & Rosch, E. (1991). *The embodied mind: Cognitive science and human experience*. MIT Press.
- Vicente, K. J. (1999). *Cognitive work analysis: Toward safe, productive, and healthy computer-based work*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Vygotsky, L. S. (1986). *Thought and language*. Cambridge, MA: The MIT Press. (Translated by A. Kozulin)
- Wan, K. (2009). A brief history of context. *International Journal of Computer Science*, 6, 33–42.
- Wason, P. C. (1960). On the failure to eliminate hypotheses in a conceptual task. *Quarterly Journal of Experimental Psychology*, 12, 129–140.
- Wason, P. C. (1966). Reasoning. In B. M. Foss (Ed.), *New horizons in psychology* (pp. 135–151). Harmondsworth, UK: Penguin Books.
- Wason, P. C. & Shapiro, D. (1971). Natural and contrived experience in a reasoning problem. *Quarterly Journal of Experimental Psychology*, 23, 63–71.
- Weyhrauch, R. W. (1980). Prolegomena to a theory of mechanized formal reasoning. *Artificial Intelligence*, 13, 133–170.
- Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin and Review*, 9, 625–636.
- Wiltgen, B. J., Sanders, M. J., Anagnostaras, S. G., Sage, J. R., & Fanselow, M. S. (2006). Context fear learning in the absence of the hippocampus. *The Journal of Neuroscience*, 26, 5484–5491.
- Winograd, T. (2001). Architectures for context. *Human-Computer Interaction*, 16, 401–419.
- Woods, D. D. (1998). Designs are hypotheses about how artifacts shape cognition and collaboration. *Ergonomics*, 41, 168–173.
- Yarbus, A. L. (1967). *Eye movements and vision*. New York, NY: Plenum Press.
- Zhang, J. & Norman, D. A. (1994). Representations in distributed cognitive tasks. *Cognitive Science*, 18, 87–122.

Paper I

Scaffolding Executive Function Capabilities via Play-&-Learn Software for Preschoolers

Anton Axelsson
Uppsala University

Richard Andersson
IT University of Copenhagen and Lund University

Agneta Gulz
Lund University and Linköping University

Educational software in the form of games or so called “computer assisted intervention” for young children has become increasingly common receiving a growing interest and support. Currently there are, for instance, more than 1,000 iPad apps tagged for preschool. Thus, it has become increasingly important to empirically investigate whether these kinds of software actually provide educational benefits for such young children. The study presented in the present article investigated whether preschoolers have the cognitive capabilities necessary to benefit from a teachable-agent-based game of which pedagogical benefits have been shown for older children. The role of executive functions in children’s attention was explored by letting 36 preschoolers (3;9–6;3 years) play a teachable-agent-based educational game and measure their capabilities to maintain focus on pedagogically relevant screen events in the presence of competing visual stimuli. Even though the participants did not succeed very well in an inhibition pretest, results showed that they nonetheless managed to inhibit distractions during game-play. It is suggested that the game context acts as a motivator that scaffolds more mature cognitive capabilities in young children than they exhibit during a noncontextual standardized test. The results further indicate gender differences in the development of these capabilities.

Keywords: inhibition, attention, teachable agents, eye tracking, learning by teaching

Through the introduction of technology in preschools, new avenues for facilitating interventions in preschool have opened up (Clements, 2002; Huffstetter, King, Onwuegbuzie, Schneider, & Powell-Smith, 2010). One important potential is the facilitation of school readiness for children who otherwise would be at risk of falling behind once they start school due to weak preparatory skills, particularly in early numeracy and literacy (Clements, Sarama, Spitzer, Lange, & Wolfe, 2011; Kendeou, van den Broek, White, & Lynch, 2007; Morgan, Farkas, & Wu, 2009). In the present study, we investigate the possibilities of introducing computer games in a revamped approach of the learning by

teaching (LBT) paradigm with the use of so called teachable agents in preschool. The LBT paradigm reverses the role of the student and lets students become teachers. However, the question is whether this kind of educational software, that has been proven pedagogically valuable for school children, is suitable for children of preschool age. In order to be able to teach, focus and attention on your tutee is crucial and this requires a sufficient development of executive control. Furthermore, the preschool is at times a distracting environment with high levels of noise and other perturbations. Thus, before investing resources in developing a full-fledged LBT-game for preschoolers and launching a longitudinal study to investigate learning effects, there are some crucial and more basic questions that need to be answered. With this study we have used a scaled down version of an LBT-game in order to investigate preschoolers’ ability to inhibit visual distractions.

This research was funded in part by the Swedish Research Council and the research environment Cognition, Communication and Learning, CCL. RA acknowledges support from the Swedish Research Council, Grant 437-2014-6735. The study has been approved by the Regional Ethical Review Board of Lund (ref. 2013/111).

Correspondence concerning this article should be addressed to Anton Axelsson, Division of Visual Information & Interaction, Department of Information Technology, PO-BOX 337, 751 05 Uppsala, Sweden. E-mail: anton.axelsson@it.uu.se

Need for Empirically Informed Educational Software Development

The impact of computer usage throughout today’s society has also affected preschool curricula in which teaching of basic technological interaction and use of computers in education is nowadays encouraged (The Swedish National Agency for Education, 2011; UNESCO, 2008). Research

on technology's impact on children's health over the past 30 years has produced divergent results. It is suggested that children in the midst of their cognitive development should have minimum technological exposure (Council on Communications and Media, 2010). In a review of neuroscientific and psychological studies related to children's exposure to digital media, Howard-Jones (2011) emphasise that we must acknowledge the factors which lead to detrimental effects on the developing brain. He concludes these factors to be (a) violent media content, (b) excessive use, and (c) late night use. Studies have shown that these factors can, for some individuals, result in attention disorders, disturbed sleep patterns, visual strain, and even seizures (Landhuis, Poulton, Welch, & Hancox, 2007; Page, Cooper, Griew, & Jago, 2010).

However, results pertaining to research on moderate use of computers and its impact on young children's learning and educational development present a more pleasant side. Children with access to computers at home during preschool age have been found to perform better on school readiness as well as motor and cognitive development tasks even when socioeconomic status is controlled for (fish, et al., 2008; Li & Atkins, 2004). Computer use in early age has also shown positive effects on language acquisition (Chera & Wood, 2003; Din & Calao, 2001), social, collaborative problem-solving (Cardelle-Elawar & Wetzel, 1995; Muller & Perlmutter, 1985), and learning motivation (Bergin, Ford, & Hess, 1993; Liu, 1996; for a review on the effects of media use on young children's learning and reasoning, see Lieberman, Bates, & So, 2009).

These mixed results leave both preschool teachers and parents struggling with how to approach the issue of letting young children interact with technology. Ljung-Djärf (2008), in a study of attitudes towards computers in three preschools in Sweden, found that there were three overall attitudes towards computer activities: (a) threatening other activities, (b) one of many alternative activities, (c) an essential activity. Preschool personnel tried their best to implement computer use in lines with the preschool curriculum. However, the choice of computer use was largely left to the child and it was mostly utilised through play separate from scheduled and structured activities.

The widespread use of computer-based technology with young children necessitates that any educational software delivers what it promises. However, the Center on Media and Child Health, USA, claims most educational video games have not been scientifically tested and thus advises parents to use their best judgement (CMCH, 2008). It is firmly believed that computers can be a valuable asset in preschool education, especially as a tool to help children who otherwise would be at risk falling behind once they start school. In order for computers to become powerful educational tools, software development must be informed by educational and developmental research on young children, and the resulting

products must be subjected to empirical investigation.

Advantages of Intervention in Preschool

Studies of school readiness have reported large individual differences among children with regard to both literacy and numeracy skills (Aunio, Hautamäki, Sajaniemi, & Van Luit, 2009; Jordan, Kaplan, Ramineni, & Locuniak, 2009). To ensure preschool children do not lag behind, it is important to consider ways to support children and help them overcome potential risks of starting school with an initial disadvantage (Denton & West, 2002; Griffin & Case, 1997; Locuniak & Jordan, 2008; Räsänen, Salminen, Wilson, Aunio, & Dehaene, 2009; Wilson, Dehaene, Dubois, & Fayol, 2009). The majority of children who enters school with early language and math difficulties are low-performers whose deficiencies stem from external factors, such as low socio-economic status (SES) and low exposure and training at home and at preschool (Denton & West, 2002; Jordan, Kaplan, Nabors Oláh, & Locuniak, 2006). Without intervention, these children are likely to remain low-performers throughout school (Jordan, Kaplan, Ramineni, & Locuniak, 2009; Kendeou, van den Broek, White, & Lynch, 2007; Mononen, Aunio, Koponen, & Aro, 2014). However, preschools are understaffed in many countries and preschool teachers often feel overloaded by what is already required from them in their everyday activities (Bullough, Hall-Kenyon, MacKay, & Marshall, 2014).

Here educational software harbours a potential with respect both to scaling-up and enabling intervention with reasonable time investment by teachers. Indeed some educational software can be used with little instruction, and teachers may be allowed to focus on one group of children while simultaneously being sure that another group of children is engaged in fun, meaningful activities whilst learning (Praet & Desoete, 2014). However, returning to a previous point, the pedagogic quality of much educational software is low. In order to benefit young children at preschool the educational software that is used must be of high quality as well as be proven pedagogically valuable for the age group in question. The study presented in this article involves a kind of educational software game proven educationally valuable for school children and investigates whether it can also be suitable for younger children.

Computer-Based Learning-by-Teaching

Educational benefits from LBT have been known since the early eighties through the seminal work of Bargh and Schul (1980). This paradigm reverses the roles by letting students become tutors in order to teach their peers. In the present paper, an explorative study is presented which investigates cognitive prerequisites in preschoolers with respect to a digital LBT game developed for this age group. The reason for this venture is that the LBT paradigm has demonstrated great

pedagogical advantages for school children. Children who take the role as tutors show an increase in effort compared to when they learn for themselves. The effort is evidenced through the children spending more time on learning materials and also by them analysing the material more thoroughly (Bargh & Schul, 1980; Martin & Schwartz, 2009). This increased effort seems to arise from motivational mechanisms (Benware & Deci, 1984). Working with learning material in order to teach others seem to bring about feelings of responsibility and meaningfulness of the task (Bargh & Schul, 1980) leading to positive effects on self-efficacy beliefs (Moore, Chang, & Smith, 2006), that is, the belief in one's own competence within a given domain. Self-efficacy beliefs in fact turn out to positively correlate with actual accomplishments (Pajares & Graham, 1999). A proposed major factor of the benefits of the LBT approach is that it stimulates metacognition (flavell, 1979), in other words, reflective thinking about problem-solving and one's own learning (Schwartz, et al., 2009).

In recent years, digital implementations of the LBT paradigm have seen light in the form of educational games involving teachable agents (TA; Brophy, Biswas, Katzberger, Bransford, & Schwartz, 1999). A TA is in essence an artificial intelligence algorithm that ensures that the behaviour of this digital representation of a tutee over time reflects how it is being taught by the human student so that the digital tutee indeed appears to learn. This form of pedagogical software, in line with research on the traditional form of LBT, has proven powerful for school children aged 8 years and upwards, both in terms of learning outcomes and motivational effects (Biswas, Leelawong, Schwartz, Vye, & The Teachable Agents Group at Vanderbilt, 2005; Ogan, et al., 2012; Pareto, Haake, Lindström, Sjödn, & Gulz, 2012).

This human-to-digital-tutee version of LBT has three unique advantages over non-digital LBT: (a) all children can be teachers, this includes those that are not naturally inclined to take such a role because they either feel less knowledgeable than their peers, or due to feelings of low self-efficacy; (b) the child who teaches can automatically be matched with the digital tutee to ensure an adequate challenge for each child tutor. To obtain this kind of match in human-to-human peer learning is often difficult due to that a large difference in competence between tutee and tutor results in non-optimal learning benefits; lastly, (c) no human tutee will suffer from a poor tutor, which can occur and be experienced as an injustice problem when LBT-inspired pedagogies are used in a group of students. The body of research that provides evidence for the educational benefits of the digital LBT approach has had a focus on pupils aged between 8 and 14 (Biswas, et al., 2005; Gulz, Haake, & Silvervarg, 2011; Kim, et al., 2006; Wagster, Tan, Wu, Biswas, & Schwartz, 2007). Whether the benefits of a digital LBT-game can be generalised to preschoolers is an open question. In particular,

the less developed executive functions in preschool children bring about doubt.

The term "executive functions" is an umbrella term for a multitude of different cognitive processes which facilitates top-down control in individuals (Diamond, 2013) and is a vital component of school readiness and academic achievement (Blair & Razza, 2007; Borella, Carretti, & Pelegrina, 2010; Zaitchik, Iqbal, & Carey, 2014). The focus of the present study was on top-down guidance or control of attention, more specifically sustained attention and inhibition. Sustained attention refers to the ability to remain alert and maintain attention on the designated task. In order to enable such focus of attention, one has to be able to suppress elements that are competing for attention; this is handled by inhibitory processes. Several researchers consider inhibition to be a primary executive control function (Burgess, Alderman, Evans, Emslie, & Wilson, 1998; Garavan, 2002; Norman & Shallice, 2000).

In order to fully benefit from LBT software that includes a digital tutee, children, in their role as teachers, must be able to pay sufficient attention to their tutee's actions and learning (Okita & Schwartz, 2013). An adequate level of attention and focus retention requires a certain developmental level with respect to executive functions, such as attentional and inhibitory capabilities. There is an intense developmental period of executive functions during preschool age (Perner & Lang, 1999) and this suggests that executive functions will not be as well developed in 3- to 6-year-olds as compared to 8-year-olds. Consequently, an educational game based upon the idea that preschool children should teach and instruct – and pay close attention to – a digital tutee may not necessarily work out well.

Although, a study by Gelman and Meck (1983) showed that children aged 3-5 were able to detect errors when a puppet performed a counting task, even when the numbers exceeded the children's explicit counting range. The study suggested that the children have implicit knowledge of numbers exceeding their apparent count limit, but due to performance demands they cannot explicate this. By observing someone else counting, the children can free up cognitive resources and therefore more easily reflect upon errors. Thus, this provides good reason for tailoring LBT-based games to preschoolers in order to alleviate cognitive strains. It is also important to emphasise that executive abilities are gradually developed (Levin, Culhane, Hartmann, Evankovich, & Mattson, 1991; Wellman & Liu, 2004).

The scientific opinion of young children's cognitive capabilities has repeatedly been revised throughout history. This is usually mediated through the introduction of novel methods and techniques, and more often than not, children turn out to be more cognitively able than previously assumed. Surprising results have been found in preschoolers' moral reasoning (Hong, 2004); infants appeal to mental

states (Baillargeon & Onishi, 2005; Southgate, Chavallier, & Csibra, 2010); and young children's selective attention and memory encoding efficacy (Blumberg & Torenberg, 2003; Markant & Amso, 2014). These results elucidate the fact that cognition does not exist in a vacuum. Especially in educational environments, skills and abilities emerge through contextual framing which acts as a scaffold for enhancing cognitive behaviour.

Digital learning games can provide this type of contextual scaffold as recently shown by Chin, Dohmen, and Schwartz (2013). Departing from Piaget's prevalent claim that 9- to 10-year-olds are not developmentally mature to reason about hierarchical relations and inheritance in taxonomies, results of their study showed that this was only true for traditional learning environments. The 9- to 10-year-olds in the study who had an opportunity to learn the same content by means of a digital game based on the LBT-pedagogy were able to reason about inheritance in taxonomies. A rich and complex digital game targets different levels of difficulty as well as different learning goals therefore it is impossible to know before empirical investigation what aspects of a game can be learnt and mastered given different developmental levels. This makes it relevant to empirically investigate to what extent 3- to 6-year-olds can have the cognitive prerequisites to pedagogically profit from LBT software.

Distractions in Preschools

The preschool environment is known to be lively with a plethora of visual and auditory distractions. In conjunction with less developed executive control in preschoolers, this might become a hindrance in introducing computer-based interventions in preschools. Visual distractions have long been known to be detrimental to preschoolers' performance on simple motor tasks (Poyntz, 1933; Somervill, Hill, White, York, & Hayes, 1978). Computers at preschools are normally situated in shared spaces where other activities are taking place; game playing might be a shared activity or other playing activities might occur around or near the child who is interacting with the computer. This implies that distractions might be of great concern especially in relation to the use of LBT-based games in preschool since players of these games need to focus on their digital tutee in order to be able to reap the benefits these games potentially have in store in terms of intervention programs in preschool.

Aim and Research Questions

Our aim in the present study was to closer examine preschoolers' distractibility by bringing an LBT-based educational game to a preschool. The following two explorative research questions were formulated:

- Are there preschoolers who can sufficiently focus on their digital tutee's actions to inhibit distractions? and if so

- How do their test scores of executive control differ from preschoolers who cannot?

Pre-tests to determine the preschoolers' sustained attention and inhibition abilities were administered. Subsequently we studied the preschoolers' inclination to be distracted and lose focus on what was central in an LBT-based game from a pedagogical design perspective. For this study, distractibility is defined as time spent gazing at pedagogically irrelevant elements within a time-limited window when focus is needed on parts relevant to the digital tutee's display of problem-solving and learning. Visual distractions were incorporated into the game in the form of animations in order to measure the effects it might have on the participants' attention. The rationale for using a game to investigate the preschoolers' level of distractibility is an ecological one with the aim to get the experiment design as representative as possible to the actual context of preschoolers interacting with a teachable agent.

Method

Participants

65 children (34 girls, 31 boys) aged 3;1 to 6;3 from a preschool in Southern Sweden were given permission through written consent forms by their guardians to participate in the experiment (70 % guardian consent rate). The particular preschool was selected because it is situated in a rural area which is representative of Sweden with regard to level of education and income among its population. In this municipality, 41 % of the inhabitants have completed higher education compared to 39 % of the population of Sweden. The average income is 298k SEK compared to 274k SEK for the average working Swede. We did not investigate any variables that might differ between families whose children were allowed to participate and families whose children were not. Although we cannot exclude the possibility that there were differences between the groups, it is thought that it may be attenuated by the nationally very small differences of SES in Sweden. The preschool houses children from ages 1 to 6 years old and the only criteria for children to participate were that they had turned 3 years of age. The study was approved by the Regional Ethical Review Board of Lund (ref. 2013/111).

Procedures and Measures

Each child participated alone in two separate data collection sessions; one pre-test session about 25 minutes long and a main test session about 15 minutes long. Data collection was carried out over a period of four weeks in April 2013; two weeks of pre-test data collection and two weeks of main test data collection. Thus there was a gap of two weeks between the two sessions for each participant. Both sessions took place in a room at the child's department of

the preschool to which the door could be closed in order to minimise uncontrollable distractions. During the pre-test sessions, the participants performed one inhibition and one sustained attention pre-test task and also played the digital LBT-game without any distractive animations in order to familiarise themselves with the game. The rationale for letting participants get familiar with the game before data collection of the main task was to make sure that we did not measure novelty effects. That is, we wanted to make sure that distractive or attentional behaviour was not induced from curiosity of the game components themselves. In the main task session, the participants played the digital LBT-game with the distracting visual stimuli.

Data collection was carried out by one experimenter who was present all through the sessions; no teachers were present during the sessions. The experimenter spent one day at the preschool prior to start of the study and was introduced to the children in order for them to feel familiar with the experimenter. The preschool served lunch at 11:30 am followed by group reading and relaxation time. All data collection sessions thus took place sometime between 10:00-11:30 and 13:00-15:00 and teachers were given the task of asking a child, who had been given parental consent, whether she or he would like to participate. Thus, no control was exerted upon time spacing between the two data collection sessions in favour for the children's individual availability and autonomy.

First pre-test: Inhibition. To measure the ability to inhibit irrelevant visual stimuli, an anti-saccade task (Hallett, 1978) embedded in a narrative to appeal to younger participants was used. The anti-saccade task is an established method of measuring inhibition of reflexive motor movements (Antoniades, et al., 2013; Hutton & Ettinger, 2006; Munoz & Everling, 2004). In this study, a narrative for the task was created in order for it to be more easily explained to the target participants; otherwise the procedure mimicked those of established tests. The task consisted of 24 trials where two apples were shown on either side of a centred diagonal cross on the screen. The participant was instructed to imagine that the apples belonged to him or her. A cartoon monster was shown to the participant and it was explained that this monster would appear and eat one of the apples, and that the only way to save the other apple was to look at it and avoid looking at the monster. Participants were asked to try and save as many apples as they could. This task was a test of the participants' inhibitory skills of reflexive motor movement when presented with visual stimuli, and is a way to measure the development of executive control with regard to inhibition. The task was presented on a computer screen and the children's eye-movements were tracked using an SMI RED remote eye tracker sampling at 250 Hz.

Children under the age of 8 have trouble suppressing reflexive saccades towards moving stimuli (Munoz & Everling,

2004). As most of the children were unlikely to pass most of the trials, it was not deemed meaningful to measure this task in terms of correct and incorrect trials. Instead the measure was calculated by using time spent avoiding looking at the monster as a fraction of the monster's display time.

Second pre-test: Sustained Attention. A traditional go-no-go paradigm task (Groot, de Sonnevill, Stins, & Boomsma, 2004; Robertson, Manly, Andrade, Baddeley, & Yiend, 1997) was adopted in order to measure sustained attention. The stimulus was presented on a computer screen and an external keyboard was used to capture the participant's response. Five colours were quasi-randomly displayed 15 times each. Each colour was displayed for 500 ms and separated by a 100 ms mask. The participant was asked to press the spacebar of the keyboard each time a new colour was shown on the screen (60 go-trials) except for when the colour was blue (15 no-go trials). Before beginning the task, all colours one at a time were displayed to the participants and they were asked to name them in order to make sure that participants were familiar with the colours and that they did not have any colour vision deficiencies that could disrupt performance. All participants correctly identified the colours. The participants were also given a test run of 15 trials after which the task began.

From the total 75 trials, each participant's final score was recorded as (a) hits, i.e., the number of times a participant withheld pressing the space bar key when the colour blue was presented; (b) misses, i.e., the number of times a participant pressed the space bar key when the colour blue was presented; (c) correct rejections, i.e., the number of times a participant pressed the space bar key when any other colour than blue was presented; and (d) false alarms, i.e., the number of times a participant withheld pressing the space bar key when any other colour than blue was presented. With these scores, a signal-detection sensitivity index — $\log d'$ — was calculated (Davison & Tustin, 1978). Participants will have an innate tendency towards being either response prone or response aversive which will lead to a biased measure if only hits are used. The calculated measure of $\log d'$ is a means to handle this response bias, and was used as the value for Sustained Attention during the analyses. Generally, d'^2 is calculated in order to handle response bias. However, $\log d'$ is recommended to use with tests of less than 100 trials (Brown & White, 2005), this since d' has a tendency to be positively biased for tests with a low number of trials (Kadlec, 1999). In order to handle extreme discriminability (i.e., a participant managing to score 100 % on either go or no-go trials), Brown and White's (2005) recommendations of adding a constant — .5 in this case — to hits, misses, correct rejections, and false alarms was adopted.

$$^1 \log d' = \frac{1}{2} \times \log_{10} \left(\frac{\text{Hits}}{\text{Misses}} \times \frac{\text{Correct Rejections}}{\text{False Alarms}} \right)$$

$$^2 d' = Z \left(\frac{\text{Hits}}{\text{Hits} + \text{Misses}} \right) - Z \left(\frac{\text{False Alarms}}{\text{False Alarms} + \text{Correct Rejections}} \right)$$

Main test: LBT-game with visually distracting stimuli.

The main task consisted of the participants playing the digital LBT-game Bird Hero – developed in JavaScript and HTML5 by Anderberg, Axelsson, Bengtsson, Håkansson, and Lindberg (2013). The game narrative revolves around a flock of chicks that are blown out of their nests and need help to get back. The child helps the chicks return home via a lift by pushing lift buttons (see screen shots in Figure 1). When a chick presents the number of feathers representing the floor it lives on, the child's task is to match this number with one of eight lift buttons presented at the bottom of the computer screen. The game consists of four game modes: (a) child plays, (b) TA watches whilst child plays, (c) child guides TA who tries to play, and (d) child watches TA play on his own. The four game modes are depicted in the bottom part of Figure 1. First, in Game Mode 1, the child alone helped the chicks to the correct branch by manoeuvring the lift panel. In Game Mode 2, the TA in the form of a panda introduced himself and asked whether he could watch in order to later on be able to help some birds himself. In Game Mode 3, the TA suggested which lift button should be pressed by presenting his choice in a thought bubble. The participant decided whether the suggestion was correct or incorrect through a binary choice by pressing a green tick or a red cross respectively. These binary buttons were presented centred at the bottom of the computer screen and the TA's thought bubble did not disappear until the participant pressed one of these binary buttons. In Game Mode 4, the TA played without any help from the participant. Participants wore headphones during game play in order to be able to listen to the TA and the birds.

It is important to emphasise that these game modes are not an experimental manipulation but a concept crucial to LBT-based games. The Bird Hero game was developed to simulate a fully working LBT-game but in a "Wizard-of-Oz"-type implementation, that is, without any advanced artificial intelligence incorporated, since we are not investigating learning effects in this particular study but instead how young children behave with a TA. This is thus, as has been expressed above, an ecological rationale.

Distraction manipulation. Throughout the game, three different distracting visual stimuli were used in the form of animations that were irrelevant to game play (Figure 2): (a) a football rolling across the grass in front of the TA and the bird, (b) an aeroplane passing by in the background, and (c) a flickering square, symbolising a program glitch. These animations were introduced experimentally to approximate the effects of a noisy environment with task-irrelevant stimuli under controlled circumstances in order to measure their influence on children's attention. Since the task-irrelevant animations were condensed into an eye-trackable area, they provided a possibility for measuring distractibility. The aim was to investigate which participants

were able to inhibit these stimuli and focus their visual attention on the task at hand. The distractive animations were played in Game Mode 3 and 4 at crucial parts of game play when the child – in order to pedagogically profit from the game – would have to concentrate on the TA.

In Game Mode 3, where the TA suggests which button to press and the child accepts or rejects the suggestion, the football rolled passed once on the lower part of the screen as the TA presented his suggestion in the thought bubble in one of the game rounds. This animation played for 3 seconds. The animation was played back when the TA made an action which the participant should attend to. Importantly, in Game Mode 3 the participant is in control of the game and can look at the thought bubble any time after the distracting animation has finished. The distraction in Game Mode 3 serves the purpose of giving a more general view of how distractions affect the participants by means of comparing two game rounds where the distracting animation is either present or absent.

In Game Mode 4, in which the child only observed the TA playing but was not able to act herself, the glitch flickered in the top left corner of the screen just as the TA made his choice on the first round (out of two). After the TA had made his choice, the aeroplane flew past diagonally, entering the top left corner of the screen. For the second round, the same two animations were played but in reversed order (i.e. aeroplane during the TA's choice and glitch after the TA's choice). These animations played for 2 seconds each.

The way the TA made his choice was by moving his hand horizontally, from left to right, along the eight lift buttons at the bottom of the computer screen. Once he reached the end of the screen, he moved his hand back from right to left and made his selection. His hand then continued all the way to the left and the hand moved horizontally once more from left to right and back again and exited the screen on the far left. The TA's hand movement across the screen took 2 seconds. The reason why the TA moves his hand along the lift buttons twice is so that when the two counterbalanced animations are played – during and after the TA's choice – the TA's hand is situated at the same spot in order to make the two conditions as visually similar as possible with the only difference that a lift button is up or down depending on whether it has been pressed by the TA or not. The animations were played 1 second before the TA reached the button he was meant to press (during the TA's choice) or had recently pressed (after the TA's choice). This is a time limited situation where the TA is in charge of the game and the child can either attend to the TA's actions or to the distracting animations – but not both. An SMI RED remote eye tracker sampling at 250 Hz was used throughout game play.

Because the game holds many moving elements – which triggers smooth-pursuit eye movements – fixations could not be reliably detected. Instead, we used gaze proportions of, or accumulated gaze time on, areas of interest (AOIs) calculated

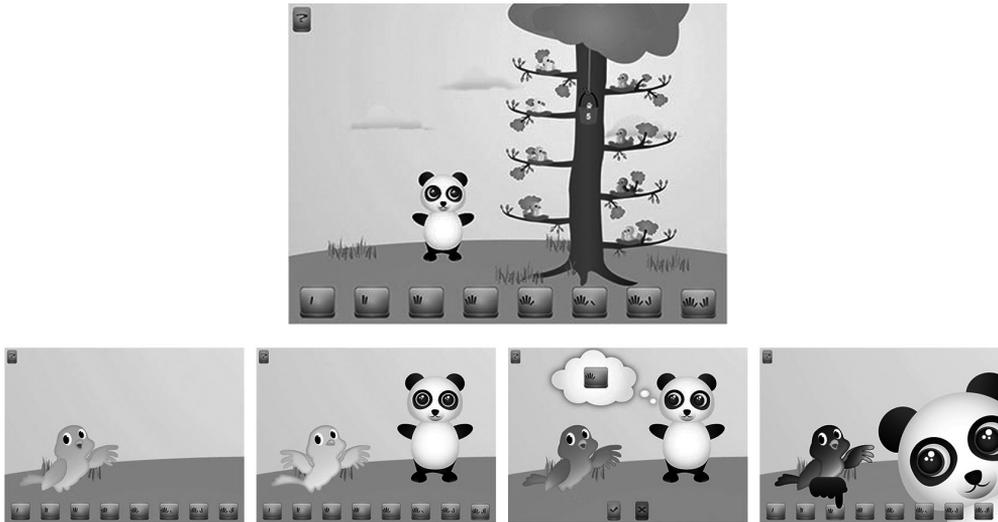


Figure 1. Screen shots of the LBT-game Bird Hero. Top: a picture of the tree with the elevator going up to the bird's nest. Bottom: the four Game Modes of the game.



Figure 2. Screen shots of the three visually distracting animations: football animation during Game Mode 3 (left); glitch (middle) and aeroplane (right) animations during Game Mode 4.

from the raw sample data. The AOIs were defined as Bird, Lift Buttons or Binary Buttons, Distraction, and TA or TA Hand. The gaze time spent on the distractive animations was used as a measure of distractibility. Comparison between animations before and after the TA's choice in Game Mode 4 gave an indication of whether children are less inclined to be distracted when the TA displays his learnt ability compared to when nothing interesting from a pedagogical perspective is happening on the screen (Research Question 1). The distractibility measure during the TA choice was used in analysis together with the pre-test measures in order to answer whether measures of executive function can predict distractibility behaviour (Research Question 2).

Results

Of the original 65 participants, 36 were part of the analysis (20 girls, 16 boys; $M_{age} = 5;2$; $SD = 9$ months). The large attrition was due to three reasons: (a) for natural reasons, a large part of the participants were not at all familiar with numbers and could therefore not participate in the main task (18 participants; $M_{age} = 4;1$); (b) a few participants were reluctant to complete all pre-tests (7 participants); and (c) the eye tracking data were too poor for some participants in the main or pre-tests (4 participants). Statistical analysis was performed using the statistical programming language R (v.2.15.1).

Table 1
Descriptive Statistics of the Study Variables

	<i>M</i>	<i>SD</i>	<i>Min</i>	<i>Max</i>
Age (years)	5.15	.74	3.76	6.25
Distractibility (ms)	198.44	260.92	0	844
Inhibition (%)	60.15	16.43	19.04	89.59
Sustained Attention (<i>log d</i>)	.46	.22	.09	.97
Hits	8.89	3.02	1	13
Misses	6.11	3.02	2	14
Correct Rejections	48.94	8.15	22	60
False Alarms	11.06	8.15	0	38

Pre-Tests Analyses

The means, standard errors, maximum and minimum values of the two pre-tests measures as well as age and Distractibility measures are summarised in Table 1. As expected, the participating preschoolers did not perform well on the inhibition task which is in line with previous research (Fukushima, Hatta, & Fukushima, 2000). On average the participants managed to completely inhibit the distraction 9 times in this task out of the 24 trials. However, using the described inhibition time fraction measure there were differences revealed across the age variable. A statistically significant positive correlation was found between age and the Inhibition measure ($r = 0.45$) whilst the correlation between age and the Sustained Attention measure, though positive, was weak ($r = 0.28$). This analysis suggests that the older a participant was, the better she or he performed on the pre-test tasks. A weak positive correlation was also found between the two pre-tests ($r = 0.29$). Student's *t*-tests were carried out and did not reveal any statistically significant difference between genders with regards to Sustained Attention ($t = 0.02$; $df = 34$; $p = 0.98$) and Inhibition ($t = -0.35$; $df = 34$; $p = 0.73$).

Distractibility Analysis

The graphs of Figure 3 show two similar time windows of the game – just when the TA presents his choice in a thought bubble of Game Mode 3 – where the difference is that the football animation was played as a distraction in the second time window (Figure 3B). In both time windows, gaze proportions are averaged over the 36 participants. Figure 3C represents a difference graph between the two time windows. On average, the participants spent 994 ms ($SE = 125$ ms) of the total 3 seconds animation playback time looking at the distraction (33 %).

The graphs of Figure 4 show the two time windows during (4A) and after (4B) the TA's choice in Game Mode 4. Gaze proportions are averaged over the 36 participants and consist of the TA helping two birds. The majority of the participants did not attend to the distracting animations at all during the TA's choice (20 out of the 36) and only 2 partic-

ipants attended to both of the animations played during the TA's choice. In Game Mode 4, the average time of which the participants gazed at the distractions during and after the TA's choice was 198 ms ($SE = 43$ ms; 9.9 % of screen time) and 581 ms ($SE = 82$ ms; 29 % of screen time) respectively. Having many participants that were not distracted lead to the data being skewed and the Distractibility measure had thus a zero-inflated distribution. To handle this, the Distractibility measure was converted to a dichotomous variable where those who were distracted (Distractibility > 0 ms) were assigned a 1 and those who were not distracted (Distractibility = 0 ms) were assigned a 0. A Yates' chi-squared test revealed a statistically significant difference in attention to the distractions during the two time windows before (16 distracted, 20 non-distracted) and after (30 distracted, 6 non-distracted) the TA choice ($X^2 = 10.17$; $df = 1$; $p < 0.01$).

Student's *t*-tests were carried out to investigate whether there were differences between those who were distracted during the TA choice from those who were not. This revealed no statistically significant differences between these two groups with regard to age or performance on the Inhibition and Sustained Attention pre-tests. However, the majority (15 of 20) of the non-distracted participants during the TA choice were female which resulted in a statistically significant Yates' chi-squared test between genders ($X^2 = 5.23$; $df = 1$; $p < 0.05$).

We used a logistic regression to analyse what pretest and participant variables could predict whether a child was distracted or not by our manipulation. The dichotomous Distractibility measure was used in the analysis against the two pre-test measures. Age and gender was also included in the analysis since age seemed to correlate with the pre-tests, and also, gender was revealed to have an impact on gaze behaviour. This analysis revealed statistically significant main effects of Sustained Attention and gender on Distractibility (Table 2). The results suggested that when controlling for age, inhibition, and sustained attention, girls were less likely distracted than boys (approximately one girl for every nine boys) by our manipulations ($b_{female} = -2.194$; $odds\ ratio = 0.111$; $p < .01$). Moreover, higher Sustained Attention pretest scores were associated with higher odds of being distracted; for every tenth of a *log d* increase of the Sustained Attention sensitivity index, the odds of being distracted increases nine-fold on average across the range ($b_{SA\ log\ d} = 4.525$; $odds\ ratio = 92$; $p < 0.05$). The pseudo R^2 for the model was 0.254 which is within range (0.2 – 0.4) of a good model fit (McFadden, 1973). A second logistic regression analysis was carried out including interactions between all predictor variables of the first model. No significant interaction effects were found.

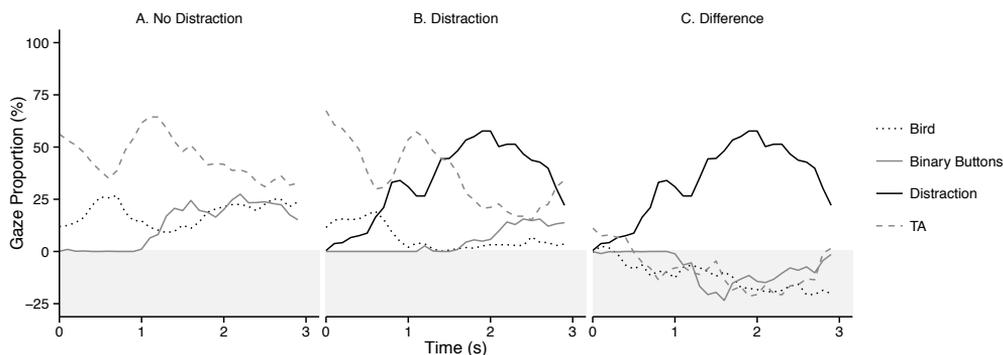


Figure 3. Gaze proportion in two similar time windows of four areas of interest over time with (A) and without (B) the football distraction in Game Mode 3. Graph C shows the resulting difference from gaze proportions of Graph A subtracted from those of Graph B. Duration is the length of the football animation distraction, and 0 on the x-axis denotes distraction onset.

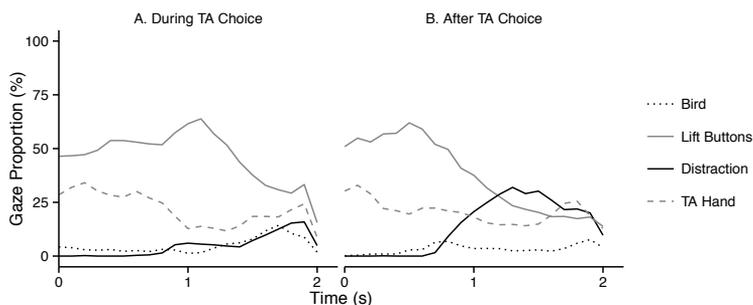


Figure 4. Gaze proportion of four areas of interest over time during (A) and after (B) TA choice in Game Mode 4. The time duration is the length of the glitch/aeroplane animation distractions, and 0 on the x-axis denotes distraction onset.

Table 2
Logistic Regression Estimates for Distractibility

	<i>b</i>	<i>SE</i>	<i>p</i>	<i>Odds ratio</i>
Intercept	-1.260	1.736	.468	—
Age (centered at <i>M</i> = 5.15)	-.460	.636	.470	.632
Gender (F)	-2.194*	.855	.010	.111
Inhibition	.158	2.821	.955	1.171
Sustained Attention (log <i>d</i>)	4.525*	2.248	.044	92.248

* *p* < 0.05.

Discussion

The pedagogical power of teachable agents (TAs) in learning environments, as a digital version of a learning-by-teaching (LBT) approach, has repeatedly been shown for students aged 8 to 14. In this study we explored the possibility of initiating the use of this kind of pedagogical software also in preschool. Due to developmental stages of executive functions it may be argued that children this young are not cognitively able to benefit from such educational games. Furthermore, a preschool is a lively environment which would further add to the doubt of whether these proposed intervention games would be suitable there. This study addressed two questions: (a) whether there are preschoolers who can inhibit distractions in order to pay attention to a TA, and (b) whether experimental measures of inhibition and sustained attention can predict the distractibility in these preschoolers.

As can be noted by the graphs in Figure 3, the distractive football animation takes quite a lot of the participants' visual attention in general when the participants are in charge of the game. Looking at the difference graph (Figure 3C) it is evident that the distraction steals equal amounts of attention from the more relevant areas of interest. This can then be contrasted with the graphs in Figure 4 which represents gaze proportions on AOIs during (4A) and after (4B) the TA makes his choice in Game Mode 4. During the TA's choice the gaze proportions of the distractive animations drop dramatically.

These results along with the presented distractibility analysis show that this group of preschoolers seem very able to inhibit distractions in order to focus their attention on their digital tutee. The participants in this study were in fact so good at this that the majority did not look at the visual distractions at all when the TA was choosing between numbers. However, after the TA made his choice the participants were once again visually occupied by the distractions as indicated by the graph in Figure 4B.

The results thus suggest that, although everything is kept constant between the two conditions, the children were more distracted after the TA had made his choice than they were during his choice selection. Interestingly, the participants in this study did not succeed well in the inhibition pre-test but nonetheless managed to inhibit during the main task performance. This shows the relevance of context and motivation in empirical investigations of cognitive capabilities. We suggest that the children's attentional behaviour is scaffolded by the context (i.e., engagement in a play-&learn software) thus they performed better in terms of inhibiting distractions than in the context of a standard inhibition test.

By inhibiting distractions, participating preschoolers could increase their attention on more important features of the game. As is shown in Graph A of Figure 4, the preschoolers does focus more of their visual attention on the TA's hand and the lift buttons, one of which their tutee is about to press,

and less on the bird and the distraction which are of less importance to benefit from the game. It is particularly interesting that the preschoolers keep such focus even though they cannot themselves be active in the game in this mode (Game Mode 4), they can only observe their tutee's actions. This result corroborates the findings of a pilot study carried out by Axelsson, Anderberg, and Haake (2013) where they found that preschoolers seem to pay attention to their TA. It also places preschoolers together with primary school children in this respect. Lindström et al. (2011) showed that primary school children paid close attention to their digital tutee whilst the tutee was acting on its own. In contrast to the preschoolers, however, the primary school children also often showed high engagement in this situation.

The present study also found similar results to those of Roderer, Krebs, Schmid, and Roebbers (2012) with regard to distractibility and engagement. In their study of selective encoding for learning, they found that preschoolers were able to increase attention towards relevant stimuli and inhibit task-irrelevant stimuli thus showing engagement in task-oriented behaviour. Roderer et al. (2012) used fairly simple and mainly static information in their study and concluded that their results were potentially dependent upon their operationalisation. However, with the results of the present study, preschoolers seem able to increase attention towards relevant stimuli also in TA-based learning environments which are more visually complex and narratively elaborated. Hence, these studies together show that preschool children are not as susceptible to visual distractions as one might believe, which further suggests that children are able to filter out distractions when their interest and focus lay elsewhere.

In regard to the second research question, the results showed that the measure of sustained attention appears to be a predictor of distractibility. Although, the results are reversed as to what one would expect. Participants that performed well on the sustained attention task were more distracted during the TA's choice. This result was surprising. One possible explanation could be related to the lack of inhibition. The children who participated in this study were shown to have poor inhibitory skills as suggested by the results of the anti-saccade task. Thus, it seems that the Sustained Attention measure captured some other aspect of attention in these participants – in relation to the distractibility measure – since motor inhibition is required also for the go-no-go paradigm task used. Our interpretation of the result is that the measure seems to have been related to the children's more general attentional abilities, that is, their tendency to notice changes in their environment. That would mean that a child who is well able to detect whenever the screen colour is blue in the sustained attention task will also be more likely to notice the visual distractions. This could suggest that overall attention to changes in the environment also leads to being more distracted unless inhibitory capabilities have matured.

Thus, when it comes to the participants that were not distracted at all, another factor must account for them being able to inhibit – or more likely filter out – the distractions.

An unanticipated result was that the female participants were less likely to attend to the distracting visual stimuli. Similar results have however been found in previous studies where boys have been found to score higher on distractibility measures (Bridges, 1929; Victor & Halverson, 1975). Poyntz (1933) found that even though boys responded to distractions more frequently, they did not spend more time being distracted than girls. Although results in the present study conversely showed that boys were on average more distracted, it is important to emphasise that the overall mean time for attention to the distracting visual stimuli during the TA choice was less than 200 ms (10 % of distraction screen time). Thus, even if a participant was distracted, regardless of gender, he or she was not distracted for long and quickly retained his or her attention to the TA. The results of a recent study of metacognitive reasoning in preschoolers showed that girls were more inclined to play another round with a TA when asked than were boys (Haake, Axelsson, Clausen-Bruun, & Gulz, 2015). A previous study (Robertson, Cross, Macleod, & Wiemer-Hastings, 2004) – including 60 somewhat older children (10-12 years old) who got to use an educational software support in either a TA or a non-TA version – showed that girls tended to interact more in the TA version than the non-TA version, whereas the pattern was reversed for the boys. Thus, there might be some motivational aspects to the digital LBT concept in general which allows girls to be slightly more focused and engaged than the boys.

Results from the present study suggest that the preschool age is the point where important cognitive capabilities for benefitting from the use of LBT games are forming. These capabilities are fairly heterogeneous in this young age group. However, the LBT game context has through this study been shown to be of practical use for scaffolding mature behaviour for some preschoolers compared to what abstract behavioural tests would suggest. Theoretically, this implies that children with underdeveloped inhibitory skills might still be able to attend to LBT-based software.

Study Limitations

Sample size. The large participant attrition in this study was not anticipated. Working with a young target population is difficult and requires a large participant marginal, so does working with eye-tracking due to difficulties in retrieving reliable data because of calibration difficulties and tracking loss. Moreover, list-wise deletion of participants unfamiliar with numbers had to be employed. This leaves the results of the present study vulnerable to only being relevant to children at the higher end of the skill spectrum of executive control. In future LBT studies with this age group, the number of participants needs to be increased. Furthermore,

if a familiarisation with numbers is required, the attrition of participants has to be estimated and accounted for in order to ensure strong statistical power. Smaller studies could consider increasing the minimum age in order to handle attrition but this will limit the generalisability of results.

Learning effects. Our research in the present study has been guided by the question whether preschoolers can profit from LBT-based games rather than do they profit. This limits us in terms of being able to say anything about learning effects with regards to preschoolers playing LBT-based games – in this case with respect to number sense and early math. However, preliminary results from a follow-up study show evidence that the LBT-based play-and-learn-game used in this study seems to have a positive impact in terms of early math learning gains (Gulz, Londos, & Haake, 2015).

Limited SES range. Since the Swedish population deviates very little in terms of SES levels, and the fact that SES levels in Sweden are fairly high, our study will have little to say about whether the cognitive prerequisites needed for LBT-based games are sufficient for children brought up in lower SES circumstances. Replications of this study of children in low SES areas as well as cross-cultural studies would be needed to draw any such conclusions.

Future Research

From the results of the present study, it seems reasonable to pursue research and development with respect to educational LBT-based software for preschoolers. The results of the study open up several future research lines. The results indicated that girls might benefit more from this pedagogical form and whether this is true must be further investigated. In any case, the display of mature cognitive behaviour of some of the preschoolers in this study shows great potential for the development of educational tools for exercising and training of preschoolers' metacognitive reasoning.

The software developed in the work of this study will be utilised as a research instrument in combination with other methods in future investigations. One objective is to find out to what extent 3- to 5-year-olds feel responsible for their tutee and at what stage the ego-protective buffer – that is, the sharing of responsibility for mistakes and errors by attributing them partly to the tutee and partly to oneself – comes into play (Chase, Chin, Oppezzo, & Schwartz, 2009). Another future objective is to investigate whether it is possible to further the development of theory of mind and metacognition in preschoolers through the use of emotional display in TAs.

Conclusions

The present study shows that the paradigm of learning by teaching implemented with teachable agent based educational games could possibly be used with much younger children than one would have thought since some of the participants in the present study possessed the prerequisites to be

able to benefit from LBT-based games. Three to six year old children who do not have mature skills at inhibiting attention to distractions can nonetheless do so when paying attention to a digital tutee they are responsible for helping. This shows that the context or task (the latter always partly defined by context or nature of the activity) influences the attentional skills of these young learners.

In conclusion, though the study suffers from some obvious limitations which affect its generalisability with regard to the results, it does show that there at least are young children that have the cognitive prerequisites to be able to play learning-by-teaching-based games. Even if not all children are able to play these games, they can be made available as soon as the child is ready. Furthermore, software games have the great potential of being individually customisable to a broader audience compared to conventional teaching methods.

References

- Anderberg, E., Axelsson, A., Bengtsson, S., Håkansson, M. & Lindberg, L. (2013). Exploring the use of a teachable agent in a mathematical computer game for preschoolers. In C. Balkenius, A. Gulz, M. Haake & B. Johansson (Eds.), *Intelligent, socially oriented technology* (pp. 161–171). Lund: Lund University Cognitive Studies, Vol 154.
- Antoniades, C., Ettinger, U., Gaymard, B., Gilchrist, I., Kristjánsson, A., Kennard, C., ... Carpenter, R. H. S. (2013). An internationally standardised antisaccade protocol. *Vision Research*, 84, 1–5.
- Aunio, P., Hautamäki, J., Sajaniemi, N. & Van Luit, J. E. H. (2009). Early numeracy in low-performing young children. *British Educational Research Journal*, 35, 25–46.
- Axelsson, A., Anderberg, E. & Haake, M. (n.d.). Can preschoolers profit from a teachable agent based play-and-learn game in mathematics? [Conference Proceedings]. In H. C. Lane, K. Yacef, J. Mostow & P. Pavlik (Eds.), *Artificial intelligence in education, Inai* (Vol. 7926, pp. 289–298). Springer.
- Baillargeon, R. & Onishi, K. H. (2005). Do 15-month-old infants understand false beliefs? *Science*, 308, 255–258.
- Bargh, J. A. & Schul, Y. (1980). On the cognitive benefits of teaching. *Journal of Educational Psychology*, 72, 583–604.
- Benware, C. A. & Deci, E. L. (1984). Quality of learning with an active versus passive motivational set. *American Educational Research Journal*, 21, 755–765.
- Biswas, G., Leelawong, K., Schwartz, D., Vye, N. & Vanderbilt, T. T. A. G. a. (2005). Learning by teaching: A new agent paradigm for educational software. *Applied Artificial Intelligence*, 19, 363–392.
- Blair, C. & Razza, R. P. (2007). Relating effortful control, executive function, and false belief understanding to emerging math and literacy ability in kindergarten. *Child Development*, 78, 647–663.
- Blumberg, F. C. & Torenberg, M. (2003). The impact of spatial cues on preschooler's selective attention. *The Journal of Genetic Psychology*, 164, 42–53.
- Borella, E., Carretti, B. & Pelegrina, S. (2010). The specific role of inhibition in reading comprehension in good and poor comprehenders. *Journal of Learning Disabilities*, 43, 541–552.
- Bridges, K. M. B. (1929). The occupational interests and attention of four-year-old children. *The Pedagogical Seminary and Journal of Genetic Psychology*, 36, 551–570.
- Brown, G. S. & White, K. G. (2005). The optimal correction for estimating extreme discriminability. *Behavior Research Methods*, 37, 436–449.
- Bullough, R. V. J., Hall-Kenyon, K. M., MacKay, K. L. & Marshall, E. E. (2014). Head start and the intensification of teaching in early childhood education. *Teaching and Teacher Education*, 37, 55–63.
- Burgess, P. W., Alderman, N., Evans, J., Emslie, H. & Wilson, B. A. (1998). The ecological validity of tests of executive function. *Journal of the International Neuropsychological Society*, 4, 547–58.
- Cardelle-Elawar, M. & Wetzel, K. (1995). Students and computers as partners in developing students' problem-solving skill. *Journal of Research on Computing in Education*, 27, 378–401.
- Chase, C. C., Chin, D. B., Oppezzo, M. A. & Schwartz, D. L. (2009). Teachable agents and the protégé effect: Increasing the effort towards learning. *Journal of Science Education and Technology*, 18, 334–352.
- Chera, P. & Wood, C. (2003). Animated multimedia 'talking books' can promote phonological awareness in children beginning to read. *Learning and Instruction*, 13, 33–52.
- Chin, D., Dohmen, I. & Schwartz, D. (2013). Young children can learn scientific reasoning with teachable agents. *IEEE Transactions on Learning Technologies*, 6, 248–257.
- Clements, D. H. (2002). Computers in early childhood mathematics. *Contemporary Issues in Early Childhood*, 160–181.
- Clements, D. H., Sarama, J., Spitler, M. E., Lange, A. A. & Wolfe, C. B. (2011). Mathematics learned by young children in an intervention based on learning trajectories: A large-scale cluster randomized trial. *Journal for Research in Mathematics Education*, 42, 127–166.
- CMCH. (2008). *Video games benefits*. Center on Media and Child Health. Retrieved from http://database.cmch.tv/mentors_parents/

video_games_benefits.asp

- Davison, M. C. & Tustin, R. D. (1978). The relation between the generalized matching law and signal-detection theory. *Journal of The Experimental Analysis of Behavior*, 29, 331-336.
- Denton, K. & West, J. (2002). *Children's reading and mathematics achievement in kindergarten and first grade* (Report). Washington, DC: National Center for Education Statistics.
- Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, 64, 135-68.
- Din, F. S. & Calao, J. (2001). The effects of playing educational video games on kindergarten achievement. *Child Study Journal*, 31, 95-102.
- Fish, A. M., Li, X., McCarrick, K., Butler, S. T., Stanton, B., Brumitt, G. A., ... Partridge, T. (2008). Early childhood computer experience and cognitive development among urban low-income preschoolers. *Journal of Educational Computing Research*, 38, 97-113.
- Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist*, 34, 906-911.
- for Education, T. S. N. A. (2011). *Curriculum for the preschool (lpfö 98)* (Vol. 2704). Stockholm, Sweden: Skolverket.
- Fukushima, J., Hatta, T. & Fukushima, K. (2000). Development of voluntary control of saccadic eye movements i. age-related changes in normal children. *Brain & Development*, 22, 173-180.
- Garavan, H. (2002). Dissociable executive functions in the dynamic control of behavior: Inhibition, error detection, and correction. *NeuroImage*, 17, 1820-1829.
- Gelman, R. & Meck, E. (1983). Preschoolers counting: Principles before skill. *Cognition*, 13, 343-359.
- Griffin, S. & Case, R. (1997). Re-thinking the primary school math curriculum: An approach based on cognitive science. *Issues in Education*, 3, 1-51.
- Groot, A., de Sonnevill, L. M. J., Stins, J. F. & Boomsma, D. I. (2004). Familial influences on sustained attention and inhibition in preschoolers. *Journal of Child Psychology and Psychiatry*, 45, 306-314.
- Gulz, A., Haake, M. & Silvervarg, A. (n.d.). Extending a teachable agent with a social conversation module – effects on student experiences and learning [Conference Proceedings]. In G. Biswas, S. Bull, J. Kay & A. Mitrovic (Eds.), *Artificial intelligence in education, Incs* (Vol. 6738, p. 106-114). Springer.
- Gulz, A., Londos, L. & Haake, M. (2015). Learning by engaged anticipation – a study of children's progress in an early math game. , Manuscript submitted for publication.
- Haake, M., Axelsson, A., Clausen-Bruun, M. & Gulz, A. (2015). Scaffolding mentalizing capabilities via a play-&-learn software for preschoolers. *Computers & Education*, Advance online publication. doi:10.1016/j.compedu.2015.09.003.
- Hallet, P. E. (1978). Primary and secondary saccades to goals defined by instructions. *Vision Research*, 18, 1279-1296.
- Hong, Y. (2004). Cultural meaning of group discussions on problematic moral situations in Korean kindergarten classrooms. *Journal of Research in Childhood Education*, 18, 179-192.
- Howard-Jones, P. (2011). *The impac of digital technologies on human wellbeing*. Oxford, UK: Nominent Trust.
- Huffstetter, M., King, J. R., Onwuegbuzie, A. J., Schneider, J. J. & Powell-Smith, K. A. (2010). Effects of a computer-based early reading program on the early reading and oral language skills of at-risk preschool children. *Journal of Education for Students Placed at Risk*, 15, 279-298.
- Hutton, S. B. & Ettinger, U. (2006). The antisaccade task as a research tool in psychopathology: A critical review. *Psychophysiology*, 46, 302-313.
- Jordan, N. C., Kaplan, D., Nabors Oláh, L. & Locuniak, M. N. (2006). Number sense growth in kindergarten: A longitudinal investigation of children at risk for mathematics difficulties. *Child Development*, 77, 153-175.
- Jordan, N. C., Kaplan, D., Ramineni, C. & Locuniak, M. N. (2009). Early math matters: Kindergarten number competence and later mathematics outcomes. *Developmental Psychology*, 45, 850-867.
- Kadlec, H. (1999). Statistical properties of d' and β estimates of signal detection theory. *Psychological Methods*, 4, 22-43.
- Kendeou, P., van den Broek, P., White, M. & Lynch, J. (2007). Preschool and early elementary comprehension: Skill development and strategy interventions. In D. S. McNamara (Ed.), *Reading comprehension strategies: Theories, interventions, and technologies* (p. 27-45). Mahwah, NJ: Erlbaum.
- Kim, S.-i., Lee, M.-J., Lee, W., So, Y., Han, C.-w., Lim, K., ... Yoon, M. (n.d.). Student modeling for adaptive teachable agent to enhance interest and comprehension [Conference Proceedings]. In *Intelligent data engineering and automated learning – ideal 2006* (Vol. 4224, p. 1234-1241). Springer.
- Landhuis, C. E., Poulton, R., Welch, D. & Hancox, R. J. (2007). Does childhood television viewing lead to attention problems in adolescence? results from a prospective longitudinal study. *Pediatrics*, 120, 532-537.
- Levin, H. S., Culhane, K. A., Hartmann, J., Evankovich, K. & Mattson, A. J. (1991). Developmental changes in performance on tests of purported frontal lobe

- functioning. *Developmental Neuropsychology*, 7, 377–395.
- Li, X. & Atkins, M. S. (2004). Early childhood computer experience and cognitive and motor development. *Pediatrics*, 113, 1715–1722.
- Lindström, P., Gulz, A., Haake, M. & Sjöden, B. (2011). Matching and mismatching between the pedagogical design principles of a math game and the actual practices of play. *Journal of Computer Assisted Learning*, 27, 90–102.
- Ljung-Djårf, A. (2008). To play or not to play — that is the question: Computer use within three Swedish preschools. *Early Education and Development*, 19, 330–339.
- Locuniak, M. N. & Jordan, N. C. (2008). Using kindergarten number sense to predict calculation fluency in second grade. *Journal of Learning Disabilities*, 41, 451–459.
- Markant, J. & Amso, D. (2014). Leveling the playing field: Attention mitigates the effects of intelligence on memory. *Cognition*, 131, 195–204.
- Martin, L. & Schwartz, D. L. (2009). Prospective adaptation in the use of representational tools. *Cognition and Instruction*, 27, 370–400.
- McFadden, D. (1973). Conditional logit analysis of qualitative choice behavior. In *Frontiers in econometrics* (p. 105–142). New York: Academic Press.
- Media, C. o. C. a. (2010). Policy statement — media education. *Pediatrics*, 126, 1–6.
- Mononen, R., Aunio, P., Koponen, T. & Aro, M. (2014). A review of early numeracy interventions for children at risk in mathematics. *International Journal of Early Childhood Special Education*, 6, 25–54.
- Moore, T. T., Chang, J. C.-J. & Smith, D. K. (2006). Clarifying the role of self-efficacy and metacognition as predictors of performance: Construct development and test. *The DATA BASE for Advances in Information Systems*, 37, 125–132.
- Morgan, P. L., Farkas, G. & Wu, Q. (2009). Five-year growth trajectories of kindergarten children with learning difficulties in mathematics. *Journal of Learning Disabilities*, 42, 306–321.
- Muller, A. A. & Perlmutter, M. (1985). Preschool children's problem-solving interactions at computers and jigsaw puzzles. *Journal of Applied Developmental Psychology*, 6, 173–186.
- Munoz, D. P. & Everling, S. (2004). Look away: The antsacade task and the voluntary control of eye movement. *Nature*, 5, 218–228.
- Norman, D. A. & Shallice, T. (2000). Attention to action: Willed and automatic control of behavior. In *Neuroscience: A reader* (p. 376–390). Malden, MA: Blackwell Publishers Inc.
- Ogan, A., Finkelstein, S., Mayfield, E., D'Adamo, C., Matsuda, N. & Cassell, J. (n.d.). "oh, dear stacy!" social interaction, elaboration, and learning with teachable agents [Conference Proceedings]. In *Proceedings of the sigchi conference on human factors in computing systems* (p. 39–48). ACM.
- Okita, S. & Schwartz, D. (2013). Learning by teaching human pupils and teachable agents: The importance of recursive feedback. *Journal of the Learning Sciences*, 22, 375–412.
- Page, A. S., Cooper, A. R., Griew, P. & Jago, R. (2010). Children's screen viewing is related to psychological difficulties irrespective of physical activity. *Pediatrics*, 126, 1011–1017.
- Pajares, F. & Graham, L. (1999). Motivation constructs, and mathematics performance. *Contemporary Educational Psychology*, 24, 124–139.
- Pareto, L., Haake, M., Lindström, P., Sjöden, B. & Gulz, A. (2012). A teachable-agent-based game affording collaboration and competition: Evaluating math comprehension and motivation. *Educational Technology Research and Development*, 60, 723–751.
- Perner, J. & Lang, B. (1999). Development of theory of mind and executive control. *Trends in Cognitive Sciences*, 3, 337–344.
- Poyntz, L. (1933). The efficacy of visual and auditory distractions for preschool children. *Child Development*, 4, 55–72.
- Praet, M. & Desoete, A. (2014). Enhancing young children's arithmetic skills through non-intensive, computerized kindergarten interventions: A randomized controlled study. *Teaching and Teacher Education*, 39, 56–65.
- Robertson, I. H., Manly, T., Andrade, J., Baddeley, B. T. & Yiend, J. (1997). 'oops': Performance correlates of everyday attentional failures in traumatic brain injured and normal subjects. *Neuropsychologia*, 35, 747–758.
- Robertson, J., Cross, B., Macleod, H. & Wiemer-Hastings, P. (2004). Children's interactions with animated agents in an intelligent tutoring system. *International Journal Of Artificial Intelligence In Education*, 14, 1560–4292.
- Roderer, T., Krebs, S., Schmid, C. & Roebers, C. M. (2012). The role of executive control of attention and selective encoding for preschoolers' learning. *Infant and Child Development*, 21, 146–159.
- Räsänen, P., Salminen, J., Wilson, A. J., Aunio, P. & Dehaene, S. (2009). Computer-assisted intervention for children with low numeracy skills. *Cognitive Development*, 24, 450–472.
- Schwartz, D. L., Chase, C., Chin, D., Oppezzo, M., Kwong, H., Okita, S., ... Biswas, G. (2009). Interactive metacognition: Monitoring and regulating a teachable agent. In *Handbook of metacognition in education* (p. 340–358). New York, NY: Routledge Press.
- Somerville, J. W., Hill, G. A., White, P. Y., York, D. C. &

- Hayes, D. M. (1978). Distractibility among advantaged and disadvantaged preschool children: Effects of cubicles vs. high levels of stimulation on task performance. *The Urban Review*, *10*, 35-44.
- Southgate, V., Chavallier, C. & Csibra, G. (2010). Seventeen-month-olds appeal to false beliefs to interpret others' referential communication. *Developmental Science*, *13*, 907-912.
- UNESCO. (2008). *Ict competency standards for teachers* (Report). Paris, France: United Nations Educational, Scientific and Cultural Organization.
- Victor, J. B. & Halverson, C. F. (1975). Distractibility and hypersensitivity: Two behavior factors in elementary school children. *Journal of Abnormal Child Psychology*, *3*, 83-94.
- Wagster, J., Tan, J., Wu, Y., Biswas, G. & Schwartz, D. (n.d.). Do learning by teaching environments with metacognitive support help students develop better learning behaviors [Conference Proceedings]. In D. S. McNamara & G. Trafton (Eds.), *Proceedings of the 29th annual meeting of the cognitive science society* (p. 695-700). Cognitive Science Society.
- Wellman, H. & Liu, D. (2004). Scaling of theory of mind tasks. *Child Development*, *75*, 523-541.
- Wilson, A. J., Dehaene, S., Dubois, O. & Fayol, M. (2009). Effects of an adaptive game intervention on accessing number sense in low-socioeconomic-status kindergarten children. *Mind, Brain, and Education*, *3*, 224-234.
- Zaitchik, D., Iqbal, Y. & Carey, S. (2014). The effect of executive function on biological reasoning in young children: An individual differences study. *Child Development*, *85*, 160-175.

Paper II

Collegial verbalisation – the value of an independent observer: an ecological approach

Anders Jansson*, Mikael Erlandsson and Anton Axelsson 

Department of Information Technology, Uppsala University, Box 337, SE-751 05, Uppsala, Sweden

(Received 31 March 2013; accepted 5 March 2015)

For knowledge elicitation in contexts where human operators are highly experienced, there are two established protocols available – concurrent and retrospective verbalisations. A third protocol – ‘conspective’ verbalisation – and a synthesised method called collegial verbalisation (CV) which combines the three protocols are presented. Where domain knowledge is shared between colleagues, one might find that they share cognitive strategies. Independent observers (colleagues) comment in the form of conspective protocols on the behaviour of target users. It solves some of the problems associated with the established verbalisation protocols. Three previously published field studies are summarised to illustrate the development of the protocol and the method, provide empirical support, and exemplify the practical value. Based on these studies, a theoretical model is presented. Conspective verbalisation is intended for use in conjunction with concurrent and retrospective verbalisations. Contributing an independent source of data is seen as the major implication of the CV method.

Keywords: verbalisation; think-aloud; expertise; cognitive ergonomics; human-factors; cognitive work analysis; strategies analysis; knowledge elicitation

1. Introduction

Research in human factors often aims for careful understanding of how users think and act in their specific work contexts and usage situations. The goal may be, for example, to account for the users’ acquired experiences while they perform different tasks since levels of expertise affect the way they interact with control systems (Rasmussen 1983, 1986). Consequently, design of control systems ought to be based on the fact that level of skill is important, for example, that strategies vary depending on level of stress (Sperandio 1978) and on ability to cope with demands in a complex and dynamic task (Jansson 1999). The fact that level of expertise is important does not, however, imply customisation of the systems to each user. On the contrary, strategies can be regarded as responses that are forced on the individual user in the form of behaviours necessary for coping with system demands, i.e. behaviour-shaping constraints (Rasmussen, Pejtersen, and Goodstein 1994; Vicente 1999). From a system development point of view, this means that it is not enough to understand *why* users behave the way they do while interacting with the system, or *what* they do in each task, but also *how* they accomplish the activities associated with a certain task or sub-task. Analyses of how activities are accomplished can be carried out through strategies analysis, either in the form of information flow maps (Rasmussen et al.

*Corresponding author. Email: anders.jansson@it.uu.se

1994; Vicente 1999), categories of cognitive response strategies (Hassall and Sanderson 2014), or heuristic strategies (Brehmer 1990; Jansson 1994, 1995). In all cases, some form of knowledge elicitation is necessary.

For the purpose of knowledge elicitation in a work context where users are highly experienced, there are several methods available. Among them are different types of verbalisation methods. Traditionally, verbal reporting is carried out through concurrent or retrospective verbalisation protocols. Both these established protocols are associated with a specific methodological challenge: there is no necessary correlation between the mental behaviour responsible for the actions taken in a certain control task and the mental behaviour behind the verbal reports about the same actions (Bainbridge 1999). This is a validity issue for both concurrent and retrospective protocols since there is no guarantee that what is verbalised is an actual account of the mental processes involved during control task performance. Historically, there were strong doubts about verbalisations as data because of this vagueness about the validity of the verbal protocols. These doubts came to an end, however, with the seminal work by Ericsson and Simon (1980, 1984) when they were able to support their claim for verbal reports as data with a strong theoretical model and concurrent verbalisations in the form of think-aloud protocols. Since then, most analysts seem to prefer concurrent verbalisation, often in the form of think-aloud procedures and protocols, over retrospective verbalisation because of the appealing theoretical support such data has (Boren and Ramey 2000). There is, however, another issue with the validity of concurrent verbalisation procedures: they may jeopardise the representativeness of the control task due to the fact that the work task is disrupted with the additional task of verbalising (e.g. Ericsson and Simon 1980, 1984; Bartl and Dörner 1998; Dickson, McLennan, and Omodei 2000). One unfortunate consequence is thus that concurrent verbalisation procedures are very difficult to apply appropriately in field studies in human factors and human–computer interaction because the participating operators cannot prioritise verbalisations without changing the way they work (Dickson et al. 2000). There are also practical issues with concurrent verbalisation procedures: it is often difficult to verbalise skill-related knowledge during task completion because much of the knowledge is tacit (Polanyi 1967). Regarding retrospective verbalisations, the challenges are even bigger. One can expect that time delays will affect the remembering of the control actions negatively in the sense of measures taken in a particular situation. The mental behaviours corresponding to these measures will decay from working memory (Gibbons 1983; Ericsson and Simon 1984). Another problem with retrospective verbalisation is that verbalisers often focus on problems closer at hand and thus infrequent problems might be overlooked (Wright and Ayton 1987). Finally, there is the problem of rationalisation. Since there is no way to separate the mental behaviour responsible for the non-observable actions taken and the mental behaviours responsible for the verbal reporting, we cannot learn from empirical data if an operator carrying out a verbalisation retrospectively is rationalising his or her behaviour (van Someren et al. 1994; Bainbridge 1999).

As a reaction to these challenges, a number of studies have explored different procedures of having others verbalise the actions of target users, for example letting colleagues or domain experts verbalise, rather than operators themselves (e.g., Jansson, Olsson, and Erlandsson 2006; Miller, Patterson, and Woods 2006; Erlandsson and Jansson 2007, 2013; McIlroy and Stanton 2011). The rationale behind the aim of having an independent observer verbalise instead of the target operator is to avoid the privacy problem (Bainbridge 1999) since this is the source to the validity issues discussed above. In this paper, we have termed this form of verbalisation ‘conspective’ due to the fact that the

verbaliser is observing whilst thinking aloud. This distinction is important because this verbalisation can be performed both in real time and with recorded material. The new protocol fits neither under concurrent verbal protocol since the verbaliser is not performing the work task, only observing it, nor under the retrospective verbal protocol since the verbaliser is seeing the events unravel for the first time.

The rationale behind concurrent, retrospective, and conspective verbal protocols is roughly similar: to extract data about mental behaviour associated with specific control task behaviour and performance. In this paper, we present a synthesised method of these three protocols called collegial verbalisation (CV). The method suggests investigators to video record target operators performing work. The critical part of the method is the conspective verbalisation where colleagues of the target operators verbalise on the recorded material. In conjunction with the conspective protocol, an investigator can choose to use (1) a concurrent protocol during recording of the target operator, (2) a retrospective protocol with the target operator, or (3) may choose to use both. Our objective is to inquire into the usefulness and value of this method by providing examples of how it was used in three field studies, and by discussing it in relation to criteria for assessing verbal reports.

1.1. Structure and aim

Since the CV method was not initially developed with reference to a theoretical framework or model, this paper is a step into the initiation of such an endeavour and it is divided into two parts. Part I is a summary of three studies where the development of the CV method is outlined. The synthesised method grew out of difficulties using established verbal protocols and focus in this part is on comparing conspective verbalisation with the established verbal protocols. The practical implications of the method and usefulness of its generated data is discussed. The overarching question investigated in the three studies is: Can colleagues verbalise on the strategies exhibited by their fellow target operators, and if they can, is it possible to generalise this ability over domains? Through empirical use and deliberation of the method, theoretical questions have been raised. Part II therefore presents the model behind the CV method. In attempting to build a theoretical framework of a verbalisation protocol, it is natural to start with Ericsson and Simon's (1984) claim on what premises such a model must be based. From these premises, five key principles for assessing verbal reports have been elaborated. In relation to these, we discuss the results of using the CV method in order to answer the question: What is the value of verbal reports based on CV or other verbalisation methods utilising knowledgeable and skilled experts as independent observers? This will help us determine whether conspective verbalisation is suitable for the purpose of verification of thought processes, or for the discovery of the same thought processes only.

2. Part I: collegial verbalisation in practice – emergence of a synthesised method

The CV method was developed iteratively and has been found to produce interesting data in our applied research projects. The data have been useful as a basis for understanding how operators work, specifically their domain-specific knowledge, often exhibited as strategies in dynamic decision-making tasks (Brehmer 1990, 1992). In the first two studies below, we describe how the method was developed in response to the inability to use the concurrent verbal protocol as the main knowledge elicitation tool. In the final study, we describe how we compared conspective against retrospective verbal protocols.

2.1. Study I: development of a new method – a field-study of train drivers

The CV method dawned in studies of train drivers. The background was the need to find methods for acquisition and assessment of their knowledge as a basis for the design of new driver interfaces in train cabs. Observational studies and interviews with train drivers showed that their domain-specific knowledge (termed route knowledge) was important for them, especially when approaching a platform and being able to stop the train at a specific position (Jansson, Olsson, and Kecklund 2005; Jansson et al. 2006). Concurrent verbal protocol procedures were applied in order to have them verbalise while they were driving. They were not prompted for explanations or descriptions of the driving task, just asked to think aloud what came into their mind while they were driving. We soon ran into problems since the train drivers went on driving without verbalising. Despite our encouragement efforts, they very often turned silent. It was particularly evident when they were riding out on the route, between two stations. Interviews with the drivers showed that they were balancing different goals in these situations; they had to keep up with the timetable, and at the same time, they tried to avoid exceeding the speed limit. They tried to drive as gentle as possible motivated by reduced energy consumption and increased comfort for the passengers. When approaching or departing from a platform they went silent for another reason: they were highly concentrated on people on the platform. To achieve safe driving in these situations, full concentration is demanded, leaving few cognitive resources free for verbalisation (Jansson et al. 2005, 2006).

Unfortunately, the drivers from the observational study were not able to participate in the retrospective knowledge elicitation phase. However, since we recorded them driving with three different cameras in the cabin, we had the opportunity to show their driving to some of their colleagues. The colleagues were asked to comment on the target drivers' driving. That is, to comment on their driving in terms of conspective verbalisations. First, this seemed to be a good idea because it was expected to yield an independent observer's opinion about each target drivers' actions. Later, it was realised that the target drivers and the colleagues were very close in experience and familiarity with the routes and the train cabs. This led to analyses on how close they were in understanding the routes, the strategies they used, and the goals they tried to achieve.

2.1.1. Method

Video recording sessions were conducted with six different professional train drivers while driving along four different types of real-schedule routes, such as long-distance routes, commuter traffic, etc. (Figure 1). They were asked to think aloud while they were driving. Seven other professional train drivers then individually performed conspective verbalisation while watching these video recordings. The recordings were muted so that they could not hear the target driver's comments. Protocol data from the target driver was used together with direct observations, interviews, and analyses from the video recordings in order to produce a model of the train driver task.

2.1.2. Summary of Study I

The analysis of the train-drivers' behaviours showed that the task of driving a train can be divided into three phases and two different kinds of actions. Table 1 shows the result of the initial analysis. This model of train driving was derived from the observational studies and the video recordings. More important, the second group of train drivers, the



Figure 1. The instrumentation and forward view of one of the studied train cabins.

colleagues, was asked to verbalise, in the same way as the target drivers had been asked to do. These comments were in accordance with the descriptions in Table 1 but were more detailed. For example, the colleagues commented that the target drivers were checking for particular signs along the track, for reference points in the surroundings on when to apply the brakes, focusing attention on people on platforms, preparing for and expecting certain braking capacity, choosing speed as a function of weighting goals such as efficiency, safety, and comfort against each other. It was concluded that the conspective protocol served as a valuable complement to other information acquisition protocols in two different ways. First, it gave a lot more pieces of information than the concurrent protocols did. Second, it specifically allowed the researchers to scrutinise the hypotheses on non-observable behaviours with the help of additional participants.

2.2. Study II: validating the new method – a case-study of operators on a high-speed ferry

The positive results of using colleagues as informants led to a decision to further the method and evaluate its usefulness in another domain. After all, the CV method was developed ad hoc as a response to the inability of the train drivers to think aloud while driving and not being available for a retrospective session. The method was formalised further, to allow for reuse, as well as scientific examination. The purpose of this study was to better understand what kind of information the conspective verbalisation could provide. Specifically, we wanted to see if the conspective protocol could provide detailed

Table 1. Examples of observable actions and non-observable actions in the three different dimensions identified in the task analyses.

	Observable actions			Non-observable actions		
Leaving a station	Controlling the platform and the doors to be shut through the mirror	Supervising, detecting, and controlling signals and signs	Listening to and watching for messages from the ATP system	Watching for people and unexpected objects	Judging time available and preparing for next section	Calculating power needed to leave station
Out on the line	Weighing speed against comfort depending on late or on time	Supervising, detecting, and controlling signals and signs	Listening to and watching for messages from the ATP system	Watching for people and unexpected objects	Judging speed ahead in order to avoid warnings or braking	Judging time in order to manage to be in time
Approaching a station	Watching for signals expected to show that the switches is clear	Supervising, detecting, and controlling signals and signs	Listening to and watching for messages from the ATP system	Watching for people and unexpected objects	Calculating braking power and braking distance needed	Preparing the entering of the next station

Note: ATP, automatic train protection.

information about inchoate cognitions, and if it could be used to scrutinise hypotheses about non-observable behaviour, in this case to detect maladaptive mental models within the high-speed craft officers. The background to this project was similar to that of the train drivers, that is, analyses of the bridges on different high-speed ferries and interviews with the crews as a basis for suggestions on how to improve the design, specifically the allocation of the new applications that were going to be part of the work environment for high-speed ferry operators (Olsson and Jansson 2006). The operators were also close colleagues in the sense that they had been driving the same ship and the same route for many years. They were highly experienced and familiar with the same routes. One of the objectives of this project was to analyse whether there were differences within the crew in terms of understanding the manoeuvring of the ship. This understanding was related to the dynamics of the ship and the surrounding environment in terms of a pre-defined route. Given the opportunity to study a high-speed ferry crew running a vessel between the mainland of Sweden and the island of Gotland, we decided to videotape the actions of the bridge crew during the entire 4-hour journey.

2.2.1. Method

Four different video cameras were used to capture the crew, instrumentation, and the surroundings (Figure 2). Two officers, one captain, and one navigator participated as target officers at the bridge during this recording. Two other officers were also present, but they did not participate in the manoeuvring of the ship. Four officers, colleagues, individually watched and verbalised the actions and decisions made by the target officers in the video. This conspective procedure was also recorded, and the protocols from these sessions were then compared to examine to what extent the four colleagues agreed on the observed behaviour. The protocols from the colleagues allowed for comparisons in-between the colleagues only, not between colleagues and target officers since we did not want to interfere with the normal procedures on the bridge by having the crew verbalising concurrently. We also knew from the train-driver study that such a procedure probably would not yield very much information.

2.2.2. Summary of Study II

The study of the high-speed ferry officers showed that the conspective protocols contributed verbal report data different from concurrent and retrospective protocols. For example, the quotes in Table 2 show how four different operators comment on the same task, navigating the ship in a fog. With the ability to compare statements in this way, inter-rater reliability data are acquired. It is a completely different kind of data source compared to the situation when there is only a single narrator available, which is the case for both



Figure 2. Four video cameras used on the high-speed ferry. From left: forward view of the ship, the captain and the first officer, instrumentation and a monitor with electronic charts, and radar integrated.

Table 2. Examples from four different protocols concerning how operators navigate when they experience fog on the route.

Officer	1	2	3	4
Comment	A lot of sea clutter (on the radar image), and many small boats. Then low visibility can be a problem, but there are quite nice features on the radar to see what is moving out there.	Fog is never amusing, and one is forced to switch to the electronic equipment, having to trust the radar instead. But I am learning to trust the electronic equipment also in good weather, so that I know how much to trust in thick fog.	In thick fog, you only see 50 metres ahead. You have to believe the radar image completely, but the small boat in front of the ship is still run over.	Fog is always unpleasant. In fog, you only have the radar, but you should still keep watch out the window. Sometimes there are disturbances in the radar, which are very similar to the echo of a real boat. That can cause some uncertainty. The radar is only a navigational aid. Aid!

concurrent and retrospective verbalisations. A detailed examination of the conspective protocols revealed that there was a high degree of agreement on the main series of events, but some specific statements conflicted between the protocols, indicating the possibility of maladaptive mental models within at least one of the narrators since both narrators' conceptions cannot be reconciled with reality at the same time (Erlandsson and Jansson 2007). It is only with the introduction of conspective protocols and the use of the synthesised CV method that it is possible to discriminate between different forms of understandings, something that can be critical in many domains, and an important input into future studies in the field of human factors.

Since the colleagues were not present at the bridge when the target officers were running the craft, they, of course, cannot remember anything from the particular situations. With the conspective verbalisation procedure, they rather recall similar situations in which they have been involved since the environmental constraints imposed on them in these situations consist in regularities that are abiding from time to time. Erlandsson and Jansson (2007) concluded that the most controversial issue with the new method is the idea of having other subjects than the target operators performing the verbalisations. With this approach, the colleagues have not been part of the target actions, and are therefore left with some form of interpretation of what they see when they verbalise. It is important to bear in mind, however, that the operators participating are highly familiar with the tasks, and that they all have long experiences from the same tasks and systems. Conspective verbalisation means a shift away from analysing working memory structures to analysing long-term memories. This also means different theoretical assumptions compared to more traditional forms of verbalisation tasks.

2.3. Study III: a quasi-experimental study of train dispatchers in a train control centre

In the first two studies, the conspective verbalisation protocol was used as a compliment to concurrent verbal protocols since we were unable to use these in the field studies. Turning to the retrospective procedure, and a comparison with the conspective procedure, we

had so far only assumed that the CV method results in more objective data than one usually gets from using only a retrospective procedure. It was therefore decided to make a systematic comparison between conspective and retrospective verbal reports. Furthermore, the overall CV method had so far been adapted and tested on operators of vehicles and vessels. Even though these domains are very different, they share some important characteristics. In both domains, the decisions made and the actions taken are based on direct perception and action, or recognition-based decisions (Klein 1993; Klein, Calderwood, and Clinton-Cirocco 2010), situations where dynamic properties are evident and apparently important. In order to investigate decisions more based on analytical problem solving, it was decided to focus on the task of supervision and control of train traffic.

2.3.1. Method

As part of an extensive research project studying train dispatchers, a systematic comparison between conspective and retrospective verbalisation was made. The verbalisation methods were used to study the effects of introducing a new software tool for planning and controlling train traffic at a train traffic control centre in Norrköping, Sweden (Andersson et al. 2014). Four professional train dispatchers were video recorded individually while working (Figure 3). These dispatchers then performed both a conspective and a retrospective verbalisation from these recordings in a quasi-experimental set-up. It made it possible to compare conspective and retrospective protocols concerning the same events.

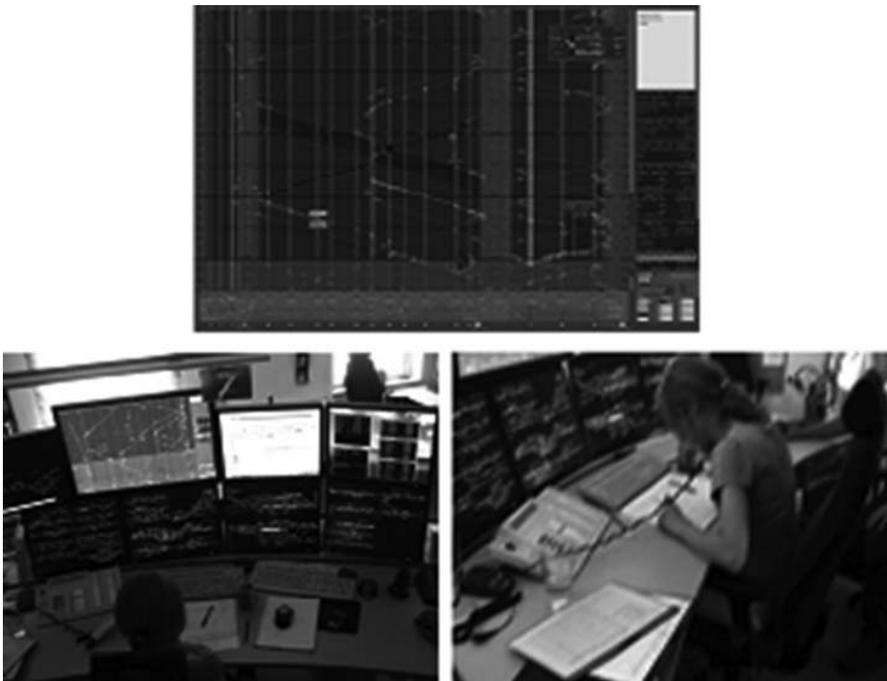


Figure 3. Bottom left: video camera capturing the monitors. Bottom right: video camera capturing operator, phone, desk, etc. Top: graphical user interface recording software capturing a specific monitor in detail.

In order to minimise the effect of remembering situation-specific details, there was a delay of a few weeks between the target situation and the two verbalisations. By this procedure, the emphasis was the comparison between long-term memory structures, assuming effects of recency to be under control.

2.3.2. Summary of Study III

Quantitative data showed that the relation between a train dispatcher who verbalised on his or her own video and a train dispatcher who verbalised on the colleague's video is quite close. Comparing the total amount of protocol data from each verbalisation shows that both conspective and retrospective verbalisation results in protocols of rather equal size (Figure 4), suggesting that colleagues can produce as many data as the person who were part of the studied events, without considering any qualitative differences.

Figure 5 shows the relation between shared and non-shared topics for each verbal protocol. It is a qualitative analysis showing that the amount of shared topics between the conspective and retrospective protocols is quite high. A Krippendorff's alpha reliability estimate showed a reliability coefficient of .8847, with a 95% confidence interval of .7912–.9627, indicating a relatively high agreement between the narrators in the two conditions retrospective and collegial verbalisation, i.e. the retrospective and conspective protocols covariate, not only with respect to shared topics, but also for non-shared topics (Erlandsson and Jansson 2013). Thus, both Figures 4 and 5 indicate, at different levels, that both narrators seem to verbalise the same content. This does not necessarily mean that they interpret all specific actions in the same way. On the contrary, for a minor part of the actions, they have different explanations of whether the actions exhibited by the target operator are relevant behaviours in the particular situation or not (Erlandsson and Jansson 2013). This information may be as important as any information showing the similarity between colleagues and target operators.

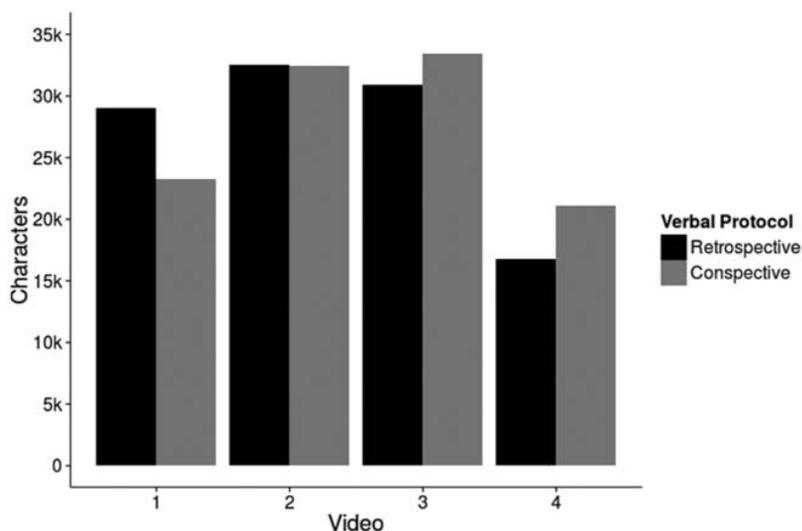


Figure 4. The amount of text in each verbal protocol, measured by the number of characters.

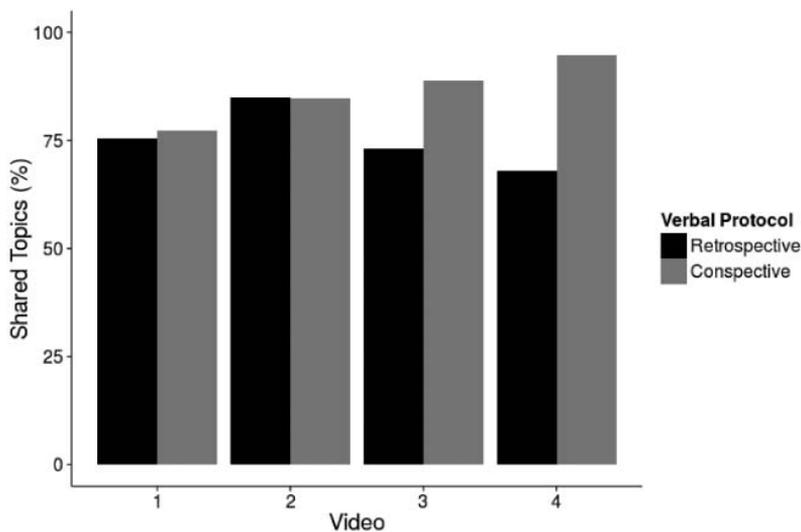


Figure 5. The relation between shared and non-shared topics for each verbal protocol.

2.4. Empirical discussion

Study I showed that the consecutive verbal protocol provided data useful for testing hypotheses about target operators' non-observable behaviours, in this case train drivers. Think-aloud statements from colleagues in the form of consecutive verbal reports allowed for interpretation of non-observable behaviours otherwise difficult to understand. Checking for particular signs along the track, for reference points in the surroundings on when to apply the brakes, focusing attention on people on platforms, preparing for and expecting certain braking capacity, choosing speed as a function of weighting different goals against each other, are all behaviours non-observable for laypersons but not for experienced colleagues. The validity of the colleagues' verbal reports seems to be a matter of inter-rater reliability, that is, judgements made by independent observers. This conclusion is quite controversial from a theoretical perspective since verbalisation procedures are supposed to shed light on mental processes associated with control actions and verbal reporting within one and the same person. In practice, however it was a matter of the number of available train drivers.

In Study II, the CV method was further developed and tested, in this case on high-speed craft operators. Again, the consecutive protocol provided data useful for testing hypotheses about target operators' non-observable behaviours. It was possible to get detailed information about inchoate cognitions. Officers on the bridge acting as target operators were silent, but their colleagues narrating on the target operators' non-observable behaviours often agreed on the types of events taking place. Here, the consecutive verbal reports contained detailed information on non-observable behaviours like checking for unexpected and irregular disturbances, planning ahead for course changes due to other ships approaching in narrow passages, and constantly looking for a room for manoeuvre. More interesting, the verbal reports also contained several examples of different interpretations on how to solve some of the situations, indicating differences in understanding. Paradoxical as it may seem, it was only with the introduction of the consecutive protocol that these differences were revealed.

Study III was designed to compare conspective verbal protocols with retrospective verbalisations, here in the case of train-traffic controllers. This was set up as a quasi-experimental design in order to study the verbal protocols at different levels of analysis. As expected, conspective and retrospective verbal protocols were similar on both protocol and topic levels. It means that the CV method is a most useful tool for mirroring the content of thoughts related to target operators' non-observable behaviours. Thus, it seems reasonable to conclude that close colleagues can verbalise on the strategies exhibited by their fellow target operators. As a consequence, it is concluded that data generated by the CV method is suitable for the purpose of verification of thought processes, and not only for the discovery of the same thought processes. The conspective protocol and the CV method have proven to be of practical value. The next central question is what value such verbal reports have from a theoretical perspective.

3. Part II: collegial verbalisation in theory – the model behind the method

3.1. Related work

The CV method is a technique primarily concerned with the process of extracting verbal report data and it can be used on different levels of analysis carried out in, for example, cognitive work analysis (Rasmussen et al. 1994; Vicente 1999). As mentioned in the introduction, the established verbal protocol methods suffer from methodological issues which originate from the problem of verification of the validity and reliability of verbal reports. Introducing an independent narrator as the one who verbalises on the actions will solve some of these problems. This has been recognised by previous authors. The elicitation by critique method (Miller et al. 2006) involves letting experts critique the actions of students or practitioners. Specifically Miller used video recording and note taking to capture the actions of a student performing an analysis task related to a rocket launch failure. The student was asked to think aloud while performing the task. This was later followed by showing screenshots and transcripts of the student's actions to a domain expert, providing cues for the expert to guide a discussion about what the expert consider correct or incorrect actions, and what they would have done themselves in the same situation. The critiquing phase was repeated for six experts individually, thereby getting multiple experts to provide information about the same topic. Similarly, the current observer narrating technique (McIlroy and Stanton 2011) involves letting an expert perform a task in a simulator, while another expert, who is close in experience to the practitioner, views the actions and narrates concurrently. Narrators are placed so that they can see the practitioner's actions and at the same time verbalise these actions without the practitioner overhearing it. This set-up gives the narrators a detailed realistic picture of the practitioner's actions. This method does not require the practitioner to think aloud, thereby avoiding the problems related to concurrent verbalisation methods. Letting a narrator who is close in experience to the practitioner verbalise the practitioner's actions rather than having an expert critiquing also provide a more detailed and realistic description about the performed actions. Another approach similar to the CV method is the stimulated recall interview procedure (Calderhead 1981). The CV method is, however, not such a procedure and the reason for this is two-fold: first, stimulated recall is used for the purpose of commentary and interpretation, but CV is thought of as a verbalisation procedure utilising two or three data generation points. In using the CV method, it is important to notice that both the verbalising target operators and their colleagues are instructed to think aloud without interrupting them with remarks for interpretations or clarifications.

They should be exposed as closely as possible to the same control task procedure. Second, one assumption behind the CV method is that environmental constraints will affect the behaviour of experienced target operators and make it possible for likewise experienced colleagues to utilise the effect of these constraints when they verbalise on the behaviour of their fellow operators. The critical part of the CV method is thus the conspective protocol, not its retrospective counterpart.

3.2. The collegial verbalisation model

The model behind the CV method consists of three data generation points: first point for data generation by a target operator (Data Generation Point 1), a second point for data generation by a colleague (Data Generation Point 2), and a third point for data generation retrospectively by the target operator (Data Generation Point 3). The three generation points are referred to as Concurrent, Conspective, and Retrospective Verbalisation, respectively (Figure 6). The first and second generation points are independent of each other in the sense that at least two different narrators are involved. The same goes for the second and the third data generation points. The separation of data generation points is seen as the unique contribution of the CV method. However, a verbal protocol from any of the generation points is not independent of the domain-specific task knowledge with which it is concerned. On the contrary, the content of the verbal reports is of central concern. Without reference to content, conspective protocols would be useless and the CV method meaningless. The consequence of this is that the method is restricted to research settings where the researcher has access to domain-specific knowledge in terms of expertise in the form of skill developed in relation to a specific task. Data generation is, of course, limited by the number of operators or colleagues that can participate.

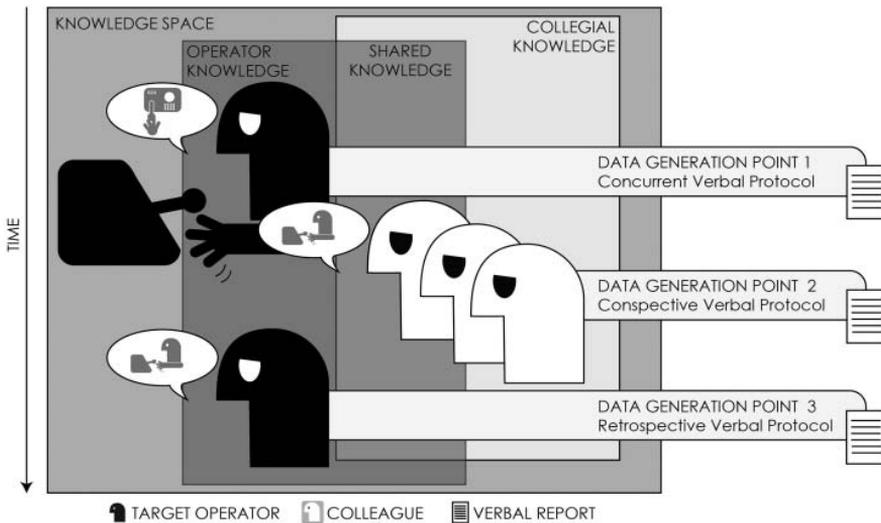


Figure 6. Illustration of the model behind the synthesised CV method with the three different data collection points: concurrent, conspective, and retrospective verbalisations leading to three different verbal protocols on the same event. Note that data collected at Generation Point 2 is intended to be used in conjunction with data from Generation Point 1 and/or Generation Point 3, not in isolation.

Even though the narrators participating are independent of each other at the data generation points, they share experiences from the same environment, which means their verbal reports will reflect these joint and common experiences. Here, it is interesting to note that Nisbett and Wilson (1977, p. 257) in their often cited review of verbalisation methods concluded that '[i]t is frightening to believe that one has no more certain knowledge of the working of one's own mind than would an outsider with intimate knowledge of one's history and of the stimuli present at the time the cognitive process occurred'. We argue that Nisbett and Wilson (1977), perhaps accidentally, pointed to two important aspects with their remark: (1) in everyday situations, verbalisations often reflect the use of both working memory and long-term memory in conjunction and (2) even though it can be hard to accept that an outsider who knows oneself well can predict one's behaviour, this points to the possibility of having other people verbalising the actions of oneself. In situations and contexts where domain knowledge is shared between close colleagues, we might find that they also share cognitive strategies.

4. Key principles for evaluating data from verbal reports

In their seminal work, Ericsson and Simon (1984, 1–8) present some issues that need to be addressed when using verbal reports. Guided by these issues, and work by other authors on this subject (Bainbridge 1999; Ericsson and Simon 1980; Ericsson and Crutcher 1991; Ericsson 2001), we propose a set of five key principles to be used in the evaluation of verbal reports made by independent observers (specifically colleagues) in interdisciplinary research settings. We first present these principles and then discuss each one in relation to the CV model presented in this paper. The first principle regards the suitability of subjects' verbalisations as scientific data (*the suitability principle*). Procedures used in verbalisation studies today seem to far from always conform to the theoretical basis invoked for the use of verbal reports as data (Boren and Ramey 2000). One important question to bear in mind when considering a verbalisation method is whether the purpose of the verbalisation data is discovering psychological processes on a general level, or whether these data are also meant to verify the result of those processes. The second principle regards the process of extracting data from behaviour (*the extraction principle*). Verbal reports are commonly used during research on human-machine interfaces in order to investigate thought processes of users. Usually, more than one data source is used in these kinds of projects. Thus, it is important to consider how one can optimise the data gathering, analysis, and interpretation processes as to make the verbalisation procedure as transparent as possible. The third principle regards the separation between data and theory (*the separation principle*). This separation is desirable in scientific research in general. Data are interpreted as objective when there is inter-subjective agreement that they correspond to the facts of the observed behaviour. In inter-disciplinary research like human-computer interaction, different techniques have evolved for recording and keeping the data in a format that makes it available for inspection, both several times and by several investigators. Thus, there is a risk that verbalised data are treated as objective, but data always have to be interpreted in relation to theory. It is therefore vital to think carefully about in what way the data-interpretation process discriminate between data and theory. The fourth principle regards the theoretical presuppositions (*the presuppositions principle*). In applied research, domain-specific knowledge is an integral part of the interpretation of the verbaliser's behaviour. This means that, contrary to laboratory settings where students are commonly used as participants and where general information processing models thus suffice, in field studies theories of long-term memory structures

becomes essential. Here, it is important to consider what the relation is between the cognitive processes in a domain-specific task and the cognitive processes responsible for generation of verbal reports regarding the same task. The fifth and final principle regards inferring thought processes from behaviour (*the inference principle*). According to Bainbridge (1999), since there is no necessary correlation between verbal reports and mental behaviour, such reports cannot be used to test theories about mental behaviour. Consequently, the resulting mental processes behind an operator's control actions may be separate from those behind the verbal report generation. It is therefore important to consider what the status is of a verbal report as evidence of a mental process.

4.1. The suitability principle

Data from any information acquisition method can, of course, be used as a source of hypotheses and predictions about any non-verbal or mental behaviour since hypotheses can have any origin (Bainbridge 1999). But as the results from our studies show, the method provides the researcher with two independent sources of data. One of these sources is the person who participated as the target user, and the other source is a close colleague. The data are independent in the sense that two observers provide two different sets of data that can be compared with each other. Thus, one source of data can be used to test predictions on task-related thinking, in the same way as other methods generating data independent from verbal reporting can be used to test these predictions. Furthermore, of the same reason, data from the CV method can also be used to test theories of verbal report generation since the two independent data sources make it possible to discriminate between task performance and verbal reporting. Thus, one overcomes the problem of confusing task knowledge with report generation as discussed by Bainbridge (1999). It, however, comes at the prize of theoretical presuppositions not yet discussed and scrutinised by the scientific community. This is discussed in [Section 4.4](#).

4.2. The extraction principle

The procedure behind the CV method, both the data-gathering part as well as the data-analysis part, need to be specified thoroughly in order to provide an as transparent description as possible of the method. We think this part can be developed further in the future. For example, issues to decide on can be whether the target user should be encouraged to think aloud or not while carrying out the task, to what extent one would like the colleague to take account of these verbal comments or not, and in what way one can measure the closeness of the colleagues' expertise. The data-analysis part need to be explicit concerning what part of the original target verbalisation that is going to be processed and analysed in order to be the task presented to the colleague.

4.3. The separation principle

In general, data are interpreted as objective when there is inter-subjective agreement that they correspond to the facts of the observed behaviour. Thus, on the data-part of this issue, the method developed here and described in this paper can be seen as science in progress since it contributes with an additional form of an independent source of data, which strengthens the situation from an inter-subjective agreement point of view. It is true both in relation to concurrent and retrospective protocols. When it comes to the theory part, the CV method can be seen as a solution to the previous difficulties of separating

verbal reporting from task-knowledge elicitation, but also as a solution to the problem of not having inter-subjective data that can be used to test theoretical assumptions regarding verbal reporting, as well as task knowledge.

4.4. *The presuppositions principle*

When verbalisation tasks are investigated without reference to information content, there is no further need to theorise than using a general information processing model, focusing on the report generation only. But when domain-specific knowledge becomes a part of the verbalisation procedure, there is evidently a need for theories, both in relation to task knowledge and the verbalisation procedure. In applied settings, researchers are often studying particular users in specific contexts of use. The domain-specific knowledge thus becomes a part of the analysis and is an important part of the interpretation of the behaviour of the verbalising subject. Often, there is a need to explain and predict behaviours of the users in the specific contexts of use. The use of the CV method shows that, if colleagues are close enough in familiarity with a specific task and system, they can verbalise strategies and other non-observable behaviours to the extent that it is possible, not only to use these report data for the purpose of general discovery of psychological processes, but also for the purpose of verification of the result of the those processes. It is perhaps the main contribution of the CV method, as well as with other methods using independent observers. This fact makes it necessary to discuss the degree of familiarity and expertise when specifying the underlying theoretical model of the verbal report generation. We believe it is necessary to develop theories on verbal reporting in relation to long-term memory structures. We propose a report generation model covering different degrees of expertise and familiarity with the task under investigation. Thus, we propose a division in terms of concurrent probing, immediate retrospective probing, retrospective probing, and finally conspectively probing. A model based on this division will make it possible to have explicit hypotheses regarding verbal reporting, including the degree of familiarity with the task. Figure 7 shows a model for predictions and hypotheses. The combination of having a target operator that can be probed concurrently or in retrospect, and a colleague that can be probed conspectively based only on the long-term memory makes it possible to combine different sorts of investigations.

Regarding task knowledge, we argue based on the results in this paper that there is a need for theoretical models for each investigated application domain. These models should specify the environmental constraints that affect the behaviour of the operators when working in each such domain. We agree with Hammond (2000) there must be theories for tasks. Such theories can be domain-specific, but they can also be semi-general, i.e. similar constraints can be aggregated over conditions, tasks, and systems.

4.5. *The inference principle*

Bainbridge (1999) discussed the process of inferring thought processes from behaviour in some length. She claimed that if data from verbal reports are used, we need to know what factors influence the way verbal reports are produced. The rationale behind this claim is that we want to minimise the distortions and maximise the validity of the evidence. The model presented here (Figure 7) allows for specifications regarding both the distortions and the claims for validity in each pair-wise combination along the time and familiarity dimensions.

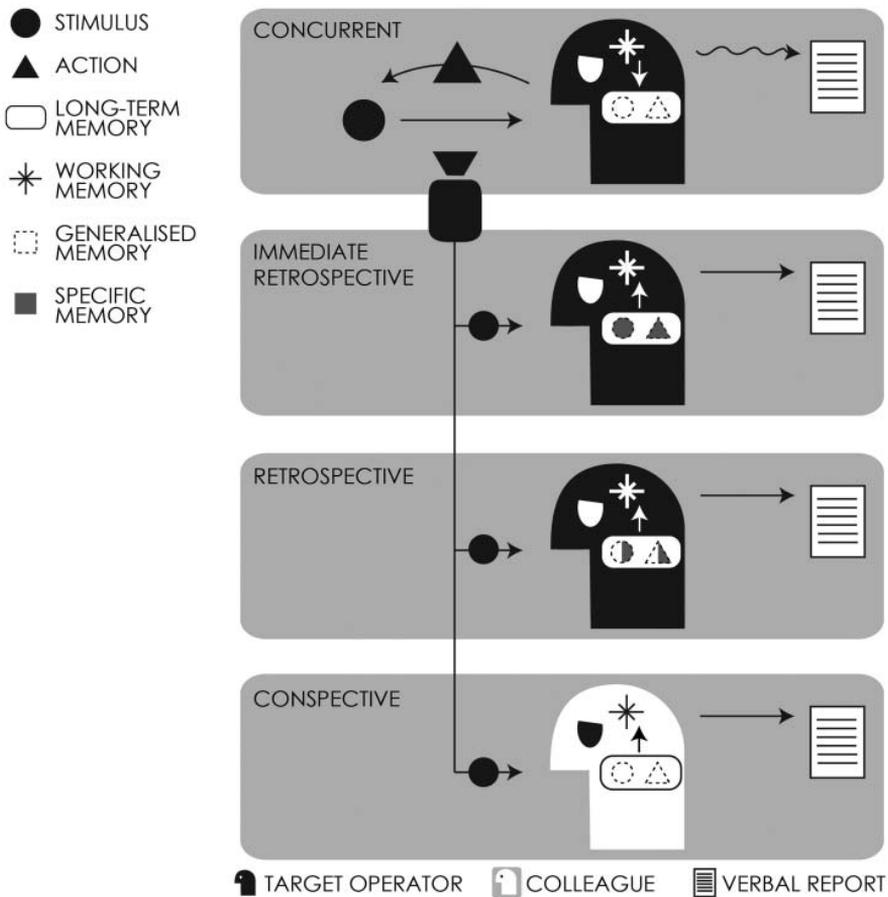


Figure 7. Illustration of verbal protocols in relation to time and familiarity for same (black) and different (white) verbaliser.

5. General discussion

We agree with Boren and Ramey (2000) that there are theoretical challenges to be solved concerning how verbal reports are used in interdisciplinary fields like human factors and cognitive ergonomics, and that Ericsson and Simon's (1984) model does not address these challenges sufficiently well. It is interesting to note that in the rather recent and thorough meta-analysis by Ericsson and co-workers (Fox, Ericsson, and Best 2011) on procedures for verbal reporting and recommendations for best reporting methods, Boren and Ramey's (2000) claim for alternative theoretical frameworks is left without any comment. It further exemplifies the lack of communication between laboratory-based and field study research. The CV method and the model behind it is an attempt to bridge this gap. But we also agree with Boren and Ramey (2000) when they claim that there is a potentially negative impact of questioning the model by Ericsson and Simon (1984) in that there are usability practitioners that may find an excuse for less rigorous procedures when gathering verbal report data. For this reason, we have emphasised that the procedure behind the

CV method should be built on verbal protocols from both target operators and colleagues so that they can be seen as complimentary data sources. In the same way, we strongly recommend that the conspective verbalisation must not be taken as a compensation and shortcut for concurrent and retrospective procedures if these techniques fail to produce analysable data or if these data are difficult to interpret. The conspective verbal protocol compliments concurrent and retrospective procedures, relying on reports from conspective protocols only does not make sense. We have outlined the assumptions behind the CV method in terms of a model describing an ecological approach to verbal report generation (i.e., Erlandsson 2014).

The model depicted in Figure 7 illustrates four types of probing. During concurrent probing ('concurrent' in Figure 7), a user receives, and acts directly on, stimuli. Through working memory, the user commits this interaction to a verbal protocol as well as to long-term memory. With immediate retrospective probing, the same user is presented with recorded stimuli of the recently performed interaction which aids remembering, that is, access to specific memories stored in long-term memory, and the user can now verbalise on the interaction. For retrospective probing, more time has passed which means that specific memories will have become integrated with general memories of similar interactions previous to, as well as after, the recorded stimuli. Lastly, in conspective probing another individual verbalise whom will try and make sense of the recorded stimuli using generalised memories of similar personal interactions. The more familiar the verbaliser is to the observed events and interactions, the more similar the verbaliser's generalised memories of such stimuli will be to the actual user. With this model, we hope to bridge the gap between laboratory-based and field study research regarding the use of verbalisation procedures. As a first step, we have in this paper addressed two of the challenges, that is, the fact that working memory and long-term memory are used in conjunction and the fact that target operators and their colleagues have shared knowledge.

Anderson (1990, 3) stated that 'we can understand a lot about human cognition without considering in detail what is inside the human head. Rather, we can look in detail at what is outside the human head and try to determine what would be optimal behaviour given the structure of the environment and the goals of the human. The claim is that we can predict behaviour of humans by assuming that they will do what is optimal. It is a different level of analysis than the analysis of mental mechanisms that has dominated information-processing psychology'. To our knowledge, no such analysis has been carried out in the field of verbal reporting. Practitioners and their colleagues develop expertise in close relation to their work tasks. It affects not only what they do in specific situations and contexts, but also how they do it. Here, we have suggested an approach for analysing and assessing how environmental constraints affect the verbal behaviour of highly experienced human operators. CV and conspective protocols can be seen as a solution to some of the problems with concurrent and retrospective protocols; the performance of the target operators will not be degraded since they do not necessarily need to verbalise, and the amount of verbalised information will be larger since the operators can go on with their more or less automated control tasks, often on a skill- or rule-based level, leaving the verbal report to a close colleague who is not busy with the control actions. The conspective protocol comes at a prize though – the colleagues who make the verbalisation have not been part of the activities, and as a consequence, cannot be assumed to have any information from a specific target situation. However, if we are interested in analysing and understanding domain-specific knowledge structures that have been developed over a long period of time, and have been used on a regular basis as responses to the demands that the specific environmental constraints impose on the operators in these situations, it may be

interesting to compare the cognitive strategies within a crew or team where all members are highly familiar with the same tasks. Switching from target operators verbalising concurrently to colleagues verbalising consecutively means also switching from information held in working memory to information recalled from long-term memory.

6. Conclusions

An ecological approach for knowledge elicitation is introduced. More specifically, a model for verbalisation by colleagues is presented as the rationale for the CV method. It is based on the idea that in situations where domain knowledge is shared between colleagues one might find that they also share cognitive strategies that they can verbalise. Independent observers (colleagues) comment in the form of consecutive verbal reports on the behaviour of a target operator. It solves some of the challenges associated with established verbalisation protocols like concurrent and retrospective verbalisation. The method is however sensitive to how close to the practitioner's experience the narrator is. Results from three previously published field studies on train drivers, high-speed ferry operators, and train dispatchers illustrate the development of the method and the new protocol, provide empirical support for this approach, and the practical value is discussed. Data generated by the CV method can, for example, be useful for practical purposes since correlations between colleagues' statements can be developed into team learning and discovery of differences between team-members understanding of situations and contexts. It makes the method particularly interesting for research in human–computer interaction, human factors, and cognitive engineering. The most controversial issue, and at the same time the unique contribution, is the fact that it is not the practitioners themselves that provide the verbalisations. The narrator is left with doing some form of interpretation of the practitioner's actions based on their knowledge and experience. CV is discussed in the light of key principles for evaluation of verbal reports, and it is concluded that consecutive protocols are separate from existing verbalisation methods but that it is intended to be used in conjunction with these, not in isolation. The major implication is the contribution of an independent source of data to be used in applied research.

Acknowledgements

The authors are grateful to the anonymous reviewers for comments helpful in improving the paper.

Disclosure statement

No potential conflict of interest was reported by the authors.

About the authors

Anders Jansson received a PhD in cognitive psychology from Uppsala University in 1997. He was appointed as an associate professor in psychology in 2006, and an associate professor in human–computer interaction in 2007. His research interests include the research fields of human factors, dynamic decision making, human–work interaction, cognitive science and human–automation interaction. He teaches human–computer interaction to graduate and undergraduate classes at Uppsala University. In his research, he collaborates with industrial partners and public organisations in the fields of train traffic control, high-speed shipping, intensive care units, train cab development and the automotive industry.

Mikael Erlandsson received a PhD in human–computer interaction from Uppsala University, where he graduated in 2014 under the supervision of Professor Anders Jansson. His research concerns methods for studying professionals in their work. In his research, he focuses on how to extract tacit knowledge and identify implicit actions. He currently works as a user research manager at FLIR Systems, where he works with usability in thermal cameras.

Anton Axelsson received an MSc in cognitive science and is currently a PhD student in computer science with a specialisation in human–computer interaction at Uppsala University, under the supervision of Professor Anders Jansson. His research focuses on human reasoning and decision making in determining how system interfaces can be designed to support professional users in making informed decisions and solving problems.

ORCID

Anton Axelsson  <http://orcid.org/0000-0002-5194-9430>

References

- Anderson, J.R. 1990. *The Adaptive Character of Thought*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Andersson, A.W., A. Jansson, B. Sandblad, and S. Tschirner. 2014. “Recognizing Complexity: Visualization for Skilled Professionals in Complex Work Situations.” In *Building Bridges: HCI, Visualization, and Non-formal Modeling*, edited by A. Ebert, G.C. van der Veer, G. Domik, N.D. Gershon, and I. Scheler, 47–66. Heidelberg: Springer.
- Bainbridge, L. 1999. “Verbal Reports as Evidence of the Process Operator’s Knowledge.” *International Journal of Human-Computer Studies* 51: 213–238.
- Bartl, C., and D. Dörner. 1998. “Speechless When Thinking: About the Role of Speech for Problem Solving.” *Sprache und Kognition* 17: 224–238.
- Boren, T., and J. Ramey. 2000. “Thinking Aloud: Reconciling Theory and Practice.” *IEEE Transactions on Professional communication* 43 (3): 261–278.
- Brehmer, B. 1990. “Strategies in Real-time, Dynamic Decision Making.” In *Insights in Decision Making*, edited by R. Hogarth, 262–279. Chicago: University of Chicago Press.
- Brehmer, B. 1992. “Dynamic Decision Making: Human Control Complex Systems.” *Acta Psychologica* 81: 211–241.
- Calderhead, J. 1981. “Stimulated Recall: A Method for Research on Teaching.” *British Journal of Educational Psychology* 51 (2): 211–217.
- Dickson, J., J. McLennan, and M.M. Omodei. 2000. “Effects of Concurrent Verbalisation on a Time-Critical, Dynamic Decision Making Task.” *Journal of General Psychology* 127: 217–228.
- Ericsson, K.A. 2001. “Protocol Analysis in Psychology.” In *International Encyclopedia of the Social and Behavioral Sciences*, edited by N. Smelser and P. Baltes, 12256–12262. Oxford: Elsevier.
- Ericsson, K.A., and R.J. Crutcher. 1991. “Introspection and Verbal Reports on Cognitive Processes – Two Approaches to the Study of Thought Processes: A Response to Howe.” *New Ideas in Psychology* 9: 57–71.
- Ericsson, K.A., and H.A. Simon. 1980. “Verbal Reports as Data.” *Psychological Review* 87: 215–251.
- Ericsson, K.A., and H.A. Simon. 1984. *Protocol Analysis: Verbal Reports as Data*. Cambridge, MA: MIT Press.
- Erlandsson, M. 2014. “What’s in Your Mind? Collegial Verbalisation: An Ecological Approach to Knowledge Elicitation.” [Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology 1147]. Uppsala: Acta Universitatis Upsaliensis.
- Erlandsson, M., and A. Jansson. 2007. “Collegial Verbalisation – A Case Study on a New Method on Information Acquisition.” *Behaviour & Information Technology* 26 (6): 535–543.
- Erlandsson, M., and A. Jansson. 2013. “Verbal Reports and Domain-specific Knowledge: A Comparison Between Collegial and Retrospective Verbalisation.” *International Journal of Cognition, Technology and Work* 15 (3): 239–254.

- Fox, M.C., K.A. Ericsson, and R. Best. 2011. "Do Procedures for Verbal Reporting of Thinking have to be Reactive? A Meta-analysis and Recommendations for Best Reporting Methods." *Psychological Bulletin* 137 (2): 316–344.
- Gibbons, F.X. 1983. "Self-attention and Self-report: The "Veridicality" Hypothesis." *Journal of Personality Assessment* 51: 517–554.
- Hammond, K.R. 2000. *Judgments Under Stress*. Oxford: Oxford University Press.
- Hassall, M., and P. Sanderson. 2014. "A Formative Approach to the Strategies Analysis Phase of Cognitive Work Analysis." *Theoretical Issues in Ergonomics Science* 15 (3): 215–261.
- Jansson, A. 1994. "Pathologies in Dynamic Decision Making: Consequences or Precursors of Failure?" *Sprache & Kognition* 13 (3): 160–173.
- Jansson, A. 1995. "Strategies in Dynamic Decision Making: Does Teaching Heuristic Strategies by Instructions Affect Performance?" In *Contributions to Decision Making*, edited by J.P. Caverni, M. Bar-Hillel, F.H. Barron, and H. Jungermann, 213–232. Amsterdam: Elsevier Science Publishers.
- Jansson, A. 1999. "Goal Achievement and Mental Models in Everyday Decision Making." In *Judgment and Decision Making: Neo-Brunswikian and Process Tracing Approaches*, edited by P. Juslin and H. Montgomery, 23–43. Mahwah, NJ: Lawrence Erlbaum Associates.
- Jansson, A., E. Olsson, and M. Erlandsson. 2006. "Bridging the Gap Between Analysis and Design: Improving Existing Driver Interfaces with Tools from the Framework of Cognitive Work Analysis." *International Journal of Cognition, Technology and Work* 8 (1): 41–49.
- Jansson, A., E. Olsson, and L. Kecklund. 2005. "Acting or Reacting? A Cognitive Work Analysis Approach to the Train Driver Task." In *Rail Human Factors: Supporting the Integrated Railway*, edited by J.R. Wilson, B. Norris, T. Clarke, and A. Mills, 40–49. Aldershot: Ashgate Publishing Limited.
- Klein, G.A. 1993. "A Recognition-primed Decision (RPD) Model of Rapid Decision Making." In *Decision Making in Action: Models and Methods*, edited by G.A. Klein, J. Orasanu, R. Calderwood, and C.E. Zsombok, 138–147. Norwood, NJ: Ablex Publishing.
- Klein, G.A., R. Calderwood, and A. Clinton-Cirocco. 2010. "Rapid Decision Making on the Fireground. The Original Study Plus a Postscript." *Journal of Cognitive Engineering and Decision Making* 4 (3): 186–209.
- McIlroy, R.C., and N.A. Stanton. 2011. "Observing the Observer: Non-intrusive Verbalisations Using the Concurrent Observer Narrative Technique." *International Journal of Cognition, Technology and Work* 13 (2): 135–149.
- Miller, J.E., E.S. Patterson, and D.D. Woods. 2006. "Elicitation by Critiquing as a Cognitive Task Analysis Methodology." *International Journal of Cognition, Technology and Work* 8 (2): 90–102.
- Nisbett, R.E., and T.D. Wilson. 1977. "Telling More than We Can Know: Verbal Reports on Mental Processes." *Psychological Review* 84: 231–259.
- Olsson, E., and A. Jansson. 2006. "Work on the Bridge – Studies of Officers on High-speed Ferries." *Behavior & Information Technology* 25 (1): 37–64.
- Polanyi, M. 1967. *The Tacit Dimension*. New York, NY: Anchor Books.
- Rasmussen, J. 1983. "Skills, Rules, and Knowledge; Signals, Signs, and Symbols, and Other Distinctions in Human Performance Models." *IEEE Transactions on Systems, Man, and Cybernetics* SMC-13: 257–266.
- Rasmussen, J. 1986. *Information Processing and Human-Machine Interaction: An Approach to Cognitive Engineering*. New York, NY: North-Holland.
- Rasmussen, J., A.M. Pejtersen, and L.P. Goodstein. 1994. *Cognitive Systems Engineering*. New York, NY: Wiley.
- Sperandio, J.C. 1978. "The Regulation of Working Methods as a Function of Workload Among Air Traffic Controllers." *Ergonomics* 21: 195–202.
- van Someren, M.W., Y.F. Barnard, and J.A.C. Sandberg. 1994. *The Think-aloud Method: A Practical Guide to Modeling Cognitive Process*. London: Academic Press.
- Vicente, K.J. 1999. *Cognitive Work Analysis: Toward Safe, Productive, and Healthy Computer-Based Work*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Wright, G., and P. Ayton. 1987. "Eliciting and Modelling Expert Knowledge." *Decision Support Systems* 3: 13–26.

Paper III

Eliciting strategies in a revolutionary domain: Exploring the hypothesis of context-free strategy categories

Ida Bodin, Anton Axelsson, and Anders Jansson
Uppsala University

Johanna Vännström
Scania CV AB

The current paper describes a strategy analysis conducted for future long haul trucks within the project Methods for Designing Future Autonomous Systems (MODAS). The MODAS project deals with a future, not yet developed system which is going to transform the work tasks and thus also the very role of the driver. In the work with the study of the present paper, the authors elaborated on a formative strategies analysis approach by having workshops with operators. The workshops were centred on scenarios familiar to the operators and which are believed to be common also in the future together with a description of the automated technology which will be available to the operators in the future. The presented procedure extends the use of the formative strategies analysis approach to also reveal implications for moving from analysis to design.

Keywords: cognitive work analysis, strategies analysis, automation, revolutionary systems design, long haul trucks

During system development it is difficult to predict how a system will be used, but it is necessary to focus on usability aspects early in the development process to lower development costs and prevent accidents. When technology advances at a rapid pace, system developers are faced with large differences between previous and potential versions of a system which further complicates development. One example is new revolutionary technical solutions for automatic control of vehicles which changes the very task of the driver. The authors of the present paper were involved in the Methods for Designing Future Autonomous Systems (MODAS; Krupenia, et al., 2014) project where one aim was to develop a human-machine interface (HMI) concept for a future long-haul vehicle with high levels of automation. Thus, automation was an a priori design attribute which revolutionises the work domain.

A first step to support the design in this circumstance can be to gain an understanding of the current work domain where higher level functions might be constant over time. But to know how to design the system in a way which do not limit users and still allow context dependent implementation of a number of strategies, an understanding of future strategies is crucial. One way to handle the unknown factors about the future system is to use pre-defined categories of strategies, as in the formative strategies analysis approach

proposed by Hassall and Sanderson (2014). The aim of the study in the present article is twofold: (1) to support the design process of a system in a revolutionary domain (i.e., a highly automated long haul truck) by conducting a formative strategies analysis; and (2) to explore how useful the approach by Hassall and Sanderson (2014) is for this purpose.

Analysis of systems in revolutionary domains

Difficulties of developing future systems are well known and this dilemma has been termed the envisioned world problem (Dekker & Woods, 1999). Difficulties usually arise because the work domain is new territory or because use of a new technology makes it difficult to predict the impact a future system might have on the work domain. A future system can either be developed from an existing system (evolutionary system) or from a clean slate (revolutionary system). In development of revolutionary systems in existing domains, methods from the Cognitive Work Analysis framework (CWA; Rasmussen, Mark Pejtersen, & Goodstein, 1994; Vicente, 1999) are very useful in describing the structure of a work domain, constraints from system purposes and its functions, the nature of tasks, strategies, and workers roles. A formative use of the CWA framework is fruitful when a new system is built from scratch (Vicente, 1999). The CWA framework can, according to Vicente, also be used in a descriptive manner in evolutionary design. However, when entering unknown territory with regard to new or drastically changing domains, CWA has its limitations since there are often no relevant work environments to model. The limits of current methodologies in relation to the development of future systems have been noted with regards to per-

Correspondence should be addressed to: Ida Bodin, Division of Visual Information & Interaction, Department of Information Technology, PO-BOX 337, 751 05 Uppsala, Sweden. E-mail: ida.bodin@it.uu.se. Telephone: +46 701679399.

formance measures (Crone, Sanderson, & Naikar, 2003), activities (Naikar & Pearce, 2003), requirements (Nehme, Scott, Cummings, & Furusho, 2006), and also in relation to revolutionary as well as intentional domains (Cummings & Guerlain, 2003).

Today, the long-haul vehicle industry is getting ever more ready to introduce fully automated trucks to our roads, that is, the technology is there to be utilised. The overall tasks of the human-machine system will not alter per se, but introduction of automation leads to dramatic shifts in task responsibilities which significantly transforms the tasks performed by human operators (Sarter, Woods, & Billings, 1997). Therefore, when technology drives the demand for new systems, there is a need to work against the idea that a clean substitution of humans with technology is feasible (substitution fallacy; *ibid.*). The idea can be counteracted by staying one step ahead and trying to anticipate transformations of the domain with regard to roles, tasks, and goals (Woods & Dekker, 2000). Furthermore, with new technology, novel strategies will emerge as operators gain experience with the system (Naikar & Pearce, 2003). With an understanding of the range of strategies that an operator could use, the system designers can build systems that support accepted user strategies in order to give some degrees of freedom for operators to grow into their work role.

Eliciting strategies in domains in the midst of a technological revolution is difficult because of the unknown cognitive strategies that will emerge from the new technology. To fulfil an ecological focus, the work domain must be stable to enable identification of strategies since these are context dependant. In working with radically different future systems, a stable domain does not exist. For the purpose of our strategies analysis of a future in-vehicle HMI, Hassall and Sanderson's (2014) formative approach thus seemed appropriate to adopt.

The Formative approach to Strategies Analysis by Hassall and Sanderson

The approach by Hassall and Sanderson (2014) contains eight strategy categories which are based on generic cognitive response strategies which, according to the authors, can be utilised in any domain. The approach is divided into two phases, the preparatory phase and the application phase.

The preparatory phase suggests researchers to identify categories of strategies through literature studies. Alternatively, researchers can use the eight categories of strategies identified through Hassall and Sanderson's (2014) own literature study based on research related to human cognition. The categories of strategies are part of a full factorial cube of three factors which exemplify when the categories of strategies are most likely to occur: (1) time pressure, (2) risk, and (3) task difficulty, which can all either be high or low. The eight categories are presented in Table 1.

The second phase (application phase) is divided into four steps:

1. Identification of activity of interest
2. Operationalization of strategy-selection criteria
3. Identification of likely categories of strategies
4. Identification of factors promoting strategy selection and change

The first step involves identifying the activities to include in the analysis, which can be a task or part of a task. The second step is identifying factors influencing likely strategies and when they occur. The time pressure, difficulties, and risk levels associated with the situations, tasks, and operators should be considered for the activities during analysis. In the third and fourth step the categories of strategies that are likely to occur are identified, and also the situations or factors prompting those strategies together with which categories of strategies that are preferred and accepted are identified.

The MODAS Project

The authors of the present paper were involved in the MODAS project with the ambition to be 'one step ahead' and investigate how the driver interaction in a highly automated truck should be designed before the technical solutions are already here. The MODAS project aimed to (1) create a method for future systems design, and (2) to apply the method to develop an in-vehicle HMI for a highly automated long-haul vehicle (Krupenia, et al., 2014). The assumptions taken within the project was that the future truck would be equipped with an automated vehicle control to help the driver handle a more complex traffic environment, with higher traffic density, higher speed limits, platoons (communicating trucks driving close together to e.g. decrease fuel consumption), and lanes that changes depending on time and day. The driver is still meant to sit in the driver seat and should be supported by the truck's interface in understanding what actions are to be implemented by the truck, as well as, for example, get information needed for route planning. These assumptions imply that future drivers are assumed to monitor the automated functions and control the vehicle on a strategic level instead of focusing on operational and tactical control (Krupenia, et al., 2014).

Using methods of CWA, the work domain for today's trucks was modelled in the beginning of the MODAS project (Bodin, 2013). When focusing on a future truck concept, there are no changes in the fact that cargo will be delivered. However, the changes in the system components and their functions as more automation is introduced, alters the strategies available in the implied sub-tasks to achieve this overall goal. Therefore there was a need within the project to investigate possible driver strategies and how to support them in the design of a future driver advisory system. The current paper describes the strategies analysis study which is based on the work domain analysis, identification of situations, and

Table 1
Strategy categories identified by Hassall and Sanderson (2014)

Strategy Category	Description	Circumstances Provoking Strategy		
		Time Pressure	Risk	Difficulty
Intuitive	Routine tasks, executed automatically	Low	Low	Low
Analytical reasoning	Used to carry out the task, based on fundamental principles of the work system, mental simulation of possible outcomes or another type of reasoning process	Low	Low	High
Compliance	Follow rules and procedures, read and follow a written procedure	Low	High	Low
Option-based	One option for action is selected (either by reasoning, evaluation or “rules-of-thumb”) from alternatives that matches some criteria	High	Low	Low
Imitation	Approach to task copied from another worker or similar situation	High	High	Low
Avoidance	Delaying or not performing the task	High	Low	High
Cue-based	Include and exclude possible actions by taking relevant evidence from the environment into account	Low	High	High
Arbitrary-choice	Tasks that are uncommon and random (scrambled, ad hoc) and without considerations of options. (Pressing a button to cancel an alarm)	High	High	High

interface concepts developed in the MODAS project.

Method

The categories of strategies identified and described by Hassall and Sanderson (2014) in their preparatory phase were used in the current study. The categories were exemplified with an initial set of suggested strategies. The reason for the initial strategy elicitation already in the preparatory phase, and without any drivers, was to be able to bring the message across to the truck drivers of the first workshop as to what we wanted to accomplish.

The application phase of the Strategies Analysis in the present study was carried out during four separate workshop events with a total of 15 long-haul truck drivers (3 female). The average age of the participants was 44 years (range: 23–59) and they had been driving trucks on an average of 20 years (range: 3–38). During the first workshop, a set of initial strategies were presented to four truck drivers in a precursory workshop for evaluation and further elicitation of strategies. In the two subsequent workshops – with four and three participating truck drivers respectively – the aim was to first discuss if the possible strategies would be acceptable and preferred and then bring the different strategies alive through presentation in interface ideas. In the last workshop session – in which four truck drivers participated – the design ideas of the previous two groups were presented. Participants were asked to evaluate the ideas of their colleagues and also asked

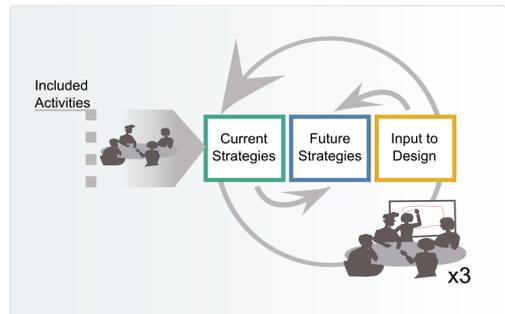


Figure 1. Driver strategies were identified from a set of initial strategies for the tasks included in the study during the initial workshop. This followed by three workshops about possible and acceptable strategies, strategy prompts and design ideas for supporting the strategies. The expressed current strategies and the input to design were also translated into future strategies during analysis.

to elaborate on them (Figure 1).

Step 1: Identification of activity of interest

In this first step, we identified the activities of interest and chose meaningful scenarios by combining the knowledge from the earlier work within the MODAS project. To identify the activities of interest in the current study a list

from Bodin (2013) of activities conducted by long haul truck drivers was used. An activity was defined as a work function occurring in a specific situation. The list included activities with work functions with many connections to higher level purposes of the abstraction hierarchy developed to describe the work domain, and with situations where the work functions were frequent. These situations were highway driving, country road driving, driving in hills, driving on slippery roads, driving in bad visibility, and start and end of a driving shift. The work functions chosen were: Cargo Transport, Short-Term Plan-Ability, Transparency of Driving Intent, Relocation, Vehicle Control (Obstacle Avoidance and Accuracy, i.e., getting to the right place), Visual Observation, and Logistics (Bodin, 2013). This list was adapted to the expected need of driver interaction in the future by removing Relocation and adding Fuel Efficient Driving. The activities Short-Term Plan-Ability, Transparency of Driving Intent, and Visual Observation were grouped together with Vehicle Control. Therefore the following functions were included in the present study:

- Vehicle Control
- Logistics
- Cargo Transport
- Fuel Efficient Driving

In the current study sub-situations or events identified as demanding by drivers were also used to bring out focus on a range of possible events, while still keeping a concrete level when discussing with drivers. The events identified as demanding by drivers were:

- Approaching traffic congestion on highway (as a threat to expected arrival time)
- Changing lanes (merging) due to road work on highway
- Passing an oncoming truck on a country road
- Following another truck during poor visibility
- Passing pedestrians during poor visibility
- Driving down a steep hill
- Driving downhill on a slippery road

A difficulty while defining the activities of interest was that the driver tasks today and in the future may not be the same. All tasks might not be performed directly by the drivers since the future truck is assumed to have a high level of automation. However, the driver's task is to make sure that these tasks are performed by the system as a whole. The activities of interest were reformulated to suit future tasks based on the expectation of the driver role to move from operational and tactical control to more holistic, strategic control. The activities included in the strategies analysis were expressed as following:

- Vehicle Control — Make sure that the truck detects everything that is important for vehicle control in the close surroundings and monitor that the truck drives efficiently. This includes observation of vehicle behaviour and monitoring of truck foresight.

- Logistics — This includes planning delivery stops and pick-ups, when and where to rest, as well as where to service the truck.

- Cargo — Transport cargo in a safe way, as well as monitoring the cargo.

- Fuel Efficient Driving — Make sure the truck is driven fuel efficiently (e.g., roll when possible).

These activities were then contextualised in the seven demanding situations presented above.

Step 2: Operationalisation of strategy-selection criteria

The second step involved evaluation of time pressure, difficulties, and risk levels associated with the task under analysis since this will affect which strategy is more likely to be chosen. Because of the uncertainties of the future demands no emphasis was put on which category of strategies that most likely would occur as suggested by Hassall and Sanderson (2014). Therefore, and because we want to support the design of a flexible system, all categories of strategies defined by Hassall and Sanderson (2014) were used.

Step 3 & 4: Identification of likely categories of strategies and factors promoting strategy selection and change

An implication of working with a future system is that there is no system to evaluate and no actual users to consult. In the beginning of the first workshop, the assumptions about the future traffic environment made within the MODAS project, and the broad concept of the future truck, were explained to the participants. To make it easier for the participants to talk about how to conduct tasks in a truck with a very high level of automation, a system they had never seen, they were asked to describe all different ways to conduct the tasks today, and then how they could be carried out in the future truck and what would prompt them. The thoughts about the future were in form of design suggestions or technical solutions for the future system.

The following three workshops started with a discussion whether the possible strategies would be acceptable, and continue with design ideas for support of the acceptable strategies. But the drivers in these workshops also provided information about strategies that could be used to perform the different tasks. Both the expressed current strategies (mostly from the first workshop) and the design suggestions were categorized into the pre-defined categories of strategies and described as future strategies during analysis.

From analysis to design – to design a system that supports the strategies

In the current study there was a need to find a way to move from the analysis of driver strategies to support the development of the future truck concept. The goal with the workshops was to get information about how to design the inter-

face, where information should be placed, how the information should be presented, and more exactly what information to present. It was important to get knowledge about how to support the acceptable strategies elicited. The approach used was similar to the approach by Olsson and Jansson (2005), where drivers worked on an interface throughout the workshop sessions. Their domain knowledge were used to find useful design metaphors and what information to display.

In the current study the last three of the four workshops were focused on how the strategies could be supported through the interface. The strategies that were conceived in the first workshop were presented during the second and third workshop. During the fourth and last workshop the ideas from the previous workshops were evaluated and elaborated upon. When the participants were given pre-defined possible strategies they were encouraged to suggest a design that support all different acceptable strategies, not only the one that is closest to how they work today or the first thing they thought of.

Seven scenario images were projected on a whiteboard during the last three workshops, illustrating the difficult situations, and were discussed in light of the four different tasks. The whiteboard had a drawing of a windshield, a tablet PC, and an instrument cluster. It was explained to the participants that the windshield would consist of a heads-up display (HUD) and that information could be presented both in the HUD, on the tablet, and in the instrument cluster. The participants were also presented with a bird's eye view and topography view, which are concepts developed within the MODAS project that shows traffic around the truck from above and a cargo delivery time line on the HUD.

Participants were asked to be creative and think of ideas on how information could be displayed that would support the strategies; this was drawn onto the whiteboard whilst switching between the seven situations that were projected on the whiteboard to be able to consider more necessary details. An example is depicted in Figure 2.

Results and Analysis

Identified likely categories of strategies and strategy selection factors

The most strategies were found for Vehicle Control and Logistics, and also large differences were seen within the analysis of those tasks, and therefore the results for those tasks are presented.

Some of the strategies used in today's trucks could be translated to strategies that could be used in the future and the design suggestions gave a hint of how the drivers thought they could solve the task in the future system which also could be translated to the strategies they would use. For example, for the task Logistics an avoidance strategy today was to ask a colleague about what route to take. The strategy of

asking a colleague could be carried out while driving a future truck as well. Because it is a relatively common strategy today, a communication tool supporting this could be developed and was suggested by the workshop participants. Another example is to use a map to plan the route, which is an analytical strategy today for the same task. When this is translated to a strategy in the future system it is generalised to 'analytical reasoning to figure out the route planning', because other tools than a map could be used in the future system. The participants also mentioned information about platoons possible to join or information about traffic accidents as needed information which would be useful if the driver is planning the route analytically.

For the Vehicle Control task, the strategies that could be used today were found to be different from those of the future system. Therefore, these strategies came mostly from suggestions for what is needed in the future as information about what is further ahead that is supporting a strategy to use environmental cues, or information about what the system is detecting which support the cue-based strategy to compare the environment with what the system detects.

The identified strategies that currently can be used and the strategies that could be used in a future MODAS system for the tasks Vehicle Control and Logistics are displayed in Table 2 and Table 3. The tables also include the strategy prompts and what category the strategies belongs to.

Design Suggestions for Support of the Accepted Strategies

The Vehicle Control Task. Cue based strategies for Vehicle Control, where environmental cues are used, can be supported by information about what comes further ahead, for example after the next turn, presented on a bird's eye view on the HUD. Another suggestion to support the cue based strategies is to show a video of what is in front of the first truck in the platoon, on the HUD, looking as if it is projected on the truck in front.

Furthermore, a cue-based strategy to compare what is observed in the environment with what is detected according to the interface, to know that the automated systems are functioning, can be supported by an interface that indicates important parts of what is detected (e.g. slow moving traffic, people or animals next to the road). It can either be information augmented at the HUD or as icons in a bird's eye view or the time and topography bar. This strategy could also be supported by information about where the truck will drive either as augmented wheel tracks or as arrows in the bird's eye view. Another suggestion is to use dotted and solid lines to show planned and implemented lane changes.

Information how the truck will handle upcoming situations on a tactic level (a little further ahead) could be shown in the bird's eye view and information about the planned

Table 2

Strategies elicited from Strategies Analysis for Vehicle Control task of both present and future vehicle systems

Strategy Category	Elicited Strategy	Present	Future	Strategy prompt
Avoidance	Use cruise control and Eco Roll	•		Highway driving
Avoidance	No supervision		•	Trust (over trust?)
Avoidance	Let passenger supervise		•	Tired driver
Intuitive	Emergency breaking/steering	•		Obstacles close in front
Intuitive	Vehicle placement in lane	•		Driving
Intuitive	Intuitively take over control		•	Emergency, Automation failure
Option-based	Either release the acceleration, hit the brakes, or swerve	•		Obstacles close in front
Option-based	Drive safe, efficient, or fast	•		Time pressure, visible police cars etc.
Option-based	Choose between alternatives of drive safe, efficient, fast or in a cargo specific mode		•	Policy, status, time pressure, cargo (hanging, liquid, livestock, or fixed)
Cue-based	Speed limit signs, GPS	•	•	Information input
Cue-based	Reads cues from traffic far in front	•	•	Visible traffic
Cue-based	Turning left/right, changing lane	•		Curves, needed lane changes
Cue-based	See where other vehicles will drive depending on how they behave (to know if the truck behaviour is sufficient)	•	•	E.g. when changing lanes in high traffic
Cue-based	The driver uses cues in the environment to drive and turn	•		Narrow place
Cue-based	Use environmental cues to know what happens further ahead	•	•	Obstacles
Cue-based	Compare environment with what the system detects		•	Driving, Wants to know that the automated systems are functioning
Cue-based	Use information from truck to know what the truck will do next		•	Lane change, Overtaking other vehicles
Compliance	Keep within speed limits	•		When driving
Compliance	The driver is following some rules of thumbs to drive and turn	•		Narrow place
Compliance	Using knowledge about how the trailer acts when turning to drive and turn	•		Narrow place
Compliance	Make sure the truck drive within speed limit		•	When driving
Compliance	Take over control		•	Temporary speed limits Traffic rules (local or terrain specific) against automation Automation failure Sensor failure
Analytical	Deciding to take over control or not by reasoning whether the truck can handle an upcoming situation		•	Novel situations

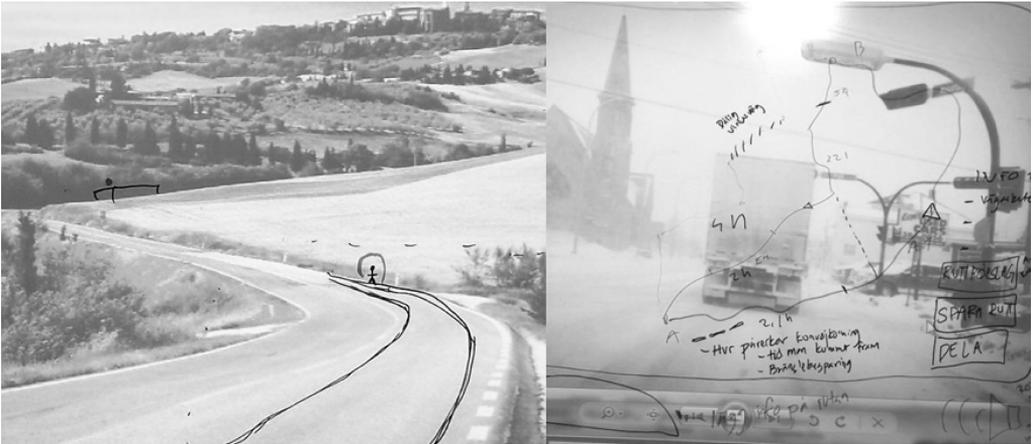


Figure 2. Picture taken of the whiteboard during one of the workshops with drivers. It is showing augmented wheel tracks on the HUD showing where the truck will drive, a detected person next to the road and a meeting truck.

route (strategic level), accidents and traffic jams could be displayed in the topography bar.

The Logistics Task. That a driver asks a colleague for directions or follows a colleague, are strategies belonging to the category for avoidance strategies but are seen as acceptable and should therefore be supported. The mentioned strategies could be used when the route is unfamiliar to the driver, maybe because of a traffic jam or other unforeseen incidents, or that the drivers want to drive together.

The strategy to ask a colleague could be supported by information about colleagues that have driven the route earlier presented on the tablet and a function to directly call up the colleague, if currently at work. The strategy to follow a colleague could be supported by a function to achieve or send your planned (or saved) routes to other drivers. In this way the driver following can have the same route as the driver in front. If the trucks have different specifications, for example different height, they might not be able to drive the same route due to a crossing flyover and then a warning should be displayed. Another idea is to show the city the truck in front is heading to on the HUD to make it easier to follow a truck that is driving in the same direction. To use a route which a colleague or the driver used before is another similar imitation strategy which can also be supported by a function to store routes which could be shared with other drivers.

The driver could have criteria for route planning that has not been foreseen and included in the route suggestions provided by the technical system. In order for the driver to better handle these situations analytical (manual) planning of the route need to be supported. Analytical planning could be supported by providing the information the technical system

base the route suggestions on directly to the driver.

The only strategy placed in the arbitrary choice category is to take the nearest possible highway exits without knowing the new route. It can be seen as an inappropriate strategy, but it was a strategy drivers described as necessary when a fast decision meant avoiding being stuck in a traffic jam. But leaving the highway can also lead to a long detour or an obstacle you cannot pass, for example a low bridge. The driver can therefore, in such a situation, be supported by information on the HUD about the shortest alternative route, which quickly gives the driver an indication whether the exit is a reasonable decision or not.

Differences between today and the future for Vehicle Control and Logistics

Many strategies for Logistics that could be used in the future system are overlapping with the strategies that can be used today, which can be seen in Table 3. When looking at Vehicle Control and Table 2 instead, many of the strategies are new and cannot be used in the present system.

Manual Vehicle Control will probably also be possible in the future, and in that way all the current strategies could also be used in the future system. The analysis of the future system was focused on the new task of monitoring the Vehicle Control rather than actually controlling the vehicle, and when the driver task changes the strategies also changes. The same division was not made for Logistics where the task was described more similar to the task today, and the driver role was assumed to be including Logistics to a higher degree, even though automated systems could be used.

Also note that the strategies described are the ones per-

Table 3
 Strategies elicited from Strategies Analysis for Logistics task of both present and future vehicle systems

Strategy Category	Elicited Strategy	Present	Future	Strategy prompt
Avoidance	Ask a colleague	•	•	Unfamiliar route, Traffic jam or other obstacles ahead
Avoidance	Follow a colleague	•	•	Shared destination, Unfamiliar route
Avoidance	Stop and wait/stay with same route even if it is a delay	•	•	Delay, no time pressure or no option
Arbitrary-choice	Take the nearest possible halting place/highway exit	•	•	When you fall behind the schedule, Missing another plan, Short time e.g. traffic jam in front on the highway, and the last exit before the jam is close
Imitation	Use colleague's route	•	•	Unfamiliar routes
Imitation	Use the same route as before	•	•	Experience of the route
Option-based	Choose from suggestions of halting places or gas stations along a route	•	•	Route planning, Unfamiliar area, Delay, Fuel level low
Option-based	Choose from alternative routes presented by system (fulfilling different criteria)	•	•	Route planning, Unfamiliar area, Delay, New delivery stop, Faulty weight values on waybill
Cue-based	Use traffic information from the radio to decide what route to take	•		When planning route
Cue-based	Read traffic signs to decide what route to take	•		When driving
Cue-based	Use traffic information and e.g. platoons that are possible to join to decide what route to take		•	When possible platoon
Compliance	Follow GPS instructions	•		When driving
Compliance	Enter truck specifications into the GPS system to ensure a sufficient route is chosen (without bridges or tunnels the truck cannot pass)		•	Company policy
Analytical	Plan the route analytically with map, GPS or road information	•	•	GPS not updated, Interested driver, wants to plan the route manually

formed by the driver, which means that some strategies are missing in the column for the future vehicle because the technology is expected to handle this task. An example of this is the avoidance strategy in the current system to use the function eco-roll which automatically controls the speed to achieve lower energy consumption. In the future when the task is to monitor the vehicle control, this is no longer a possible avoidance strategy.

Discussion

As mentioned by Vicente (1999), flexibility is needed in an open system and can be achieved by letting the worker adapt the task procedure to the current situation, rather than

enforcing one 'correct' way to perform a task through system design. To give the workers the flexibility needed to control an open system was the rational for identifying many acceptable strategies to be supported with system design in the current study. The categories of strategies and the approach suggested by Hassall and Sanderson (2014) were useful because it facilitated a focus on possible strategies instead of focusing on (the best) strategies used today. It was a way to ask the drivers for a variety of possible strategies, and by considering all categories more strategies seemed to be found.

By adopting the context-independent approach with predefined categories suggested by Hassall and Sanderson (2014), it was possible to avoid a too heavy focus on the

work domain of today. A limited focus on today's domain was necessary to support the development of the future HMI, which is assumed to change heavily and be very different from the system today.

The categories of strategies suggested by Hassall and Sanderson

The categories of strategies by Hassall and Sanderson are very generic, which means that they are not developed for a specific work domain but for many possible actions. The method differs heavily in this sense from the ecological focus encouraged by Vicente (1999). As mentioned earlier, the context-free approach seems to be an advantage when working with revolutionary domains. But what needs to be noted here is that the focus is on the strategies for the tasks implemented by the driver in the future system. By studying the affordances and constraints put on the work from the assumed future traffic situation and physical world it might be possible to instead describe the strategies the future system (truck and driver) as a whole can use. We did not investigate how useful the approach by Hassall and Sanderson is when system strategies are of interest. The aim with the current study was to identify many possible future strategies that the driver could implement to enable us to better understand information and design requirements. Because of this, the method by Hassall and Sanderson (2014) was used as is without changing or excluding any of the suggested categories.

During the data collection and analysis within the current study, conducted for the truck domain, it was noted that some of the categories of strategies were used to a larger extent. Some other categories of strategies were not seen to be applicable to the tasks conducted by truck drivers (see Table 2 and 3). For example, no strategy in the current study was described as an intuition strategy. Some of the strategies described could be placed in that category, but when other categories were a better fit the intuition strategy category was seen as superfluous. This implies that the proposed categories of strategies by Hassall and Sanderson (2014) are overlapping, which is further exemplified by the fact that, for example, all strategies elicited could also be placed in the 'imitation' category, if learned by copying another driver or if used before. That the categories are overlapping seem to come from that the categories are describing how to discover strategies, and therefore one specific strategy could belong to different categories; it can be identified in different ways. Furthermore, a strategy from the category for option-based strategies needs to be complemented with another strategy for the selection of one of the options. The category for option-based strategies was within the current study used for strategies including a selection of an action or plan suggested by a technical system. It is then necessary to further consider what strategy to use to select one alternative, and how the selecting strategy can be supported.

Limitations

The main limitation with the method by Hassall and Sanderson (2014), and our adaptation of their approach, is the difficulty to validate the results. Today we cannot determine how the analysis will influence the design and final system, or the success of the system because it is long until the future system will be built and can be used in real situations.

The approach by Hassall and Sanderson also included a cube containing different factors which determined when a strategy from the different categories was more likely to be used. The factors are (1) time pressure, (2) risk, and (3) task difficulty, which can be either low or high. The cube was of limited use in the current study for two reasons. First, the current study aimed at identifying many possible strategies to support interface and system design with the aim to give the driver flexibility when coping with different situations by letting the driver chose how to conduct the task when in the situation. Second, it was hard to determine the levels for the cubes three factors when working with a revolutionary system, especially because the levels are context dependent.

Something else to note is that we did not identify any strategies that should be avoided and thus be made impossible to perform in the new system design. An explanation of why no strategies that should be rejected were found is that the study was focused on identifying how future driver tasks can be performed, and how technology can support the driver in performing the tasks with different strategies. In the truck domain it was difficult to imagine obvious bad strategies, and if proposed they would probably immediately be rejected. It could be because the actions implemented often gives a direct visible impact in the environment (e.g. steering), and therefore bad strategies often seem very obvious, as opposed to other domains with more indirect feedback such as process control.

It is also difficult to say that a strategy never would be useful or optimal for a particular situation. One example is the strategy in the Arbitrary Choice category of taking the nearest possible highway exit without knowing if it is appropriate. The drivers discussed that information could support the driver to not end up in the situation of being uninformed, they further argued that the strategy will somebullet be needed also in the future because of too short time to consider the information (e.g., between the occurrence of a traffic jam and the last possible exit before it).

Conclusions

Strategies that would be acceptable to use in the new system of the revolutionary domain under analysis were identified through workshops with today's workers. The strategies and design suggestions from the workers were presented to the MODAS interface designer as an input to development of the future truck concept. It is difficult to evaluate the actual

impact of the current study on the final design, and it is many years until we will be able to see the concept as an end product. The formative approach for strategies analysis proposed by Hassall and Sanderson (2014) was useful in identifying many possible strategies that could be used by the worker to perform tasks in a future system. The approach could be especially useful for development of work environments in domains with large technological advances and the environment in which the system operates is drastically changing. This because their method can be used even when the system is not yet developed, and the study is based on early concepts and assumptions about the future.

Acknowledgments

The authors would like to thank the drivers from Scania Transport Laboratory that participated in this study. The work presented in this manuscript has been funded by the Strategic Vehicle Research and Innovation (FFI) funding scheme to the MODAS project (2012-03678).

References

- Bodin, I. (2013). *Using cognitive work analysis to identify opportunities for enhancing human-heavy vehicle system performance*. Stockholm, Sweden. (Unpublished Master's Thesis)
- Cornelissen, M., Salmon, P. M., Jenkins, D. P. & Lenné, M. G. (2011). How can they do it? a structured approach to the strategies analysis phase of cognitive work analysis. In *Proceedings of the human factors and ergonomics society annual meeting* (p. 340-344). Human Factors and Ergonomics Society.
- Crone, D. J., Sanderson, P. M. & Naikar, N. (2003). Using cognitive work analysis to develop a capability for the evaluation of future systems. In *Proceedings of the human factors and ergonomics society 47th annual meeting* (p. 1938-1942). Thousand Oaks, CA: SAGE Publications.
- Cummings, M. L. & Guerlain, S. (2003). The tactical tomahawk conundrum: designing decision support systems for revolutionary domains. In *Smc'03 conference proceedings* (p. 1583-1588). IEEE: 2003 IEEE International Conference on Systems, Man and Cybernetics.
- Dekker, S. W. & Woods, D. D. (1999). To intervene or not to intervene: the dilemma of management by exception. *Cognition, Technology and Work, 1*, 86-96.
- Hassall, M. E. & Sanderson, P. M. (2014). A formative approach to the strategies analysis phase of cognitive work analysis. *Theoretical Issues in Ergonomics Science, 15*, 1-47.
- Hassall, M. E., Sanderson, P. M. & Cameron, I. T. (2014). The development and testing of safer: A resilience-based human factors method. *Journal of Cognitive Engineering and Decision Making, 8*, 162-186.
- Jenkins, D. P., Stanton, N. A., Salmon, P. M. & Guy, W. H. (2009). *Cognitive work analysis: Coping with complexity*. Farnham, UK: Ashgate Publishing, Ltd.
- Krupenia, S., Selmarker, A., Fagerlönn, J., Delsing, K., Jansson, A., Sandblad, B. & Grane, C. (2014). The methods for designing future autonomous systems (modas) project. In T. Ahram, W. Karwowski & T. Marek (Eds.), *Proceedings of the 5th international conference on applied human factors and ergonomics*. Proceedings of the 5th International Conference on Applied Human Factors and Ergonomics.
- Naikar, N., Moylan, A. & Pearce, B. (2006). Analysing activity in complex systems with cognitive work analysis: Concepts, guidelines and case study for control task analysis. *Theoretical Issues in Ergonomics Science, 7*, 371-394.
- Naikar, N. & Pearce, B. (2003). Analysing activity for future systems. In *Proceedings of the human factors and ergonomics society 47th annual meeting* (p. 1928-1932). Thousand Oaks, CA: SAGE Publications.
- Nehme, C. E., Scott, S. D., Cummings, M. L. & Furusho, C. Y. (2006). Generating requirements for futuristic heterogenous unmanned systems. In *Proceedings of the human factors and ergonomics society 50th annual meeting* (p. 235-239). Thousand Oaks, CA: SAGE Publications.
- Olsson, E. & Jansson, A. (2005). Participatory design with train drivers: A process analysis. *Interacting with Computers, 17*, 147-166.
- Rasmussen, J., Mark Pejtersen, A. & Goodstein, L. (1994). *Cognitive systems engineering*. New York, NY: Wiley.
- Roth, E. M. (2009). Understanding cognitive strategies for shared situation awareness across a distributed system: An example of strategies analysis. In A. M. Bisantz & C. M. Burns (Eds.), *Applications of cognitive work analysis* (p. 129-148). Boca Raton, FL: CRS Press.
- Sarter, N. B., Woods, D. D. & Billings, C. E. (1997). Automation failures. In G. Salvendy (Ed.), *Handbook of human factors & ergonomics* (p. 1926-1943). New York: Wiley.
- Vicente, K. J. (1999). *Cognitive work analysis: Toward safe, productive, and healthy computer-based work*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Woods, D. & Dekker, S. (2000). Anticipating the effects of technological change: A new era of dynamics for human factors. *Theoretical Issues in Ergonomics Science, 1*, 272-282.
- Woods, D. D., Johannesen, L. J., Cook, R. I. & Sarter, N. B. (1994). *Behind human error: Cognitive systems, computers, and hindsight*. Dayton, OH: CSERIAC.

Recent licentiate theses from the Department of Information Technology

- 2016-005** Ida Bodin: *Cognitive Work Analysis in Practice: Adaptation to Project Scope and Industrial Context*
- 2016-004** Kasun Hewage: *Towards a Secure Synchronous Communication Architecture for Low-power Wireless Networks*
- 2016-003** Sven-Erik Ekström: *A Vertex-Centered Discontinuous Galerkin Method for Flow Problems*
- 2016-002** Rubén Cubo: *Mathematical Modeling for Optimization of Deep Brain Stimulation*
- 2016-001** Victor Shcherbakov: *Radial Basis Function Methods for Pricing Multi-Asset Options*
- 2015-006** Hanna Holmgren: *Towards Accurate Modeling of Moving Contact Lines*
- 2015-005** Siyang Wang: *Analysis of Boundary and Interface Closures for Finite Difference Methods for the Wave Equation*
- 2015-004** Pavol Bauer: *Parallelism and Efficiency in Discrete-Event Simulation*
- 2015-003** Fredrik Hellman: *Multiscale and Multilevel Methods for Porous Media Flow Problems*
- 2015-002** Ali Dorostkar: *Developments in preconditioned iterative methods with application to glacial isostatic adjustment models*
- 2015-001** Karl Ljungkvist: *Techniques for Finite Element Methods on Modern Processors*
- 2014-007** Ramūnas Gutkovas: *Advancing Concurrent System Verification: Type based approach and tools*



UPPSALA
UNIVERSITET

Department of Information Technology, Uppsala University, Sweden